OFELIA

ICT-258365

Deliverable D4.1

First year report on planning development testing and operation of individual islands

Editor:	*Wouter Tavernier, IBBT*
Work Package (leader)	WP4, IBBT
Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	M13
Actual delivery date:	M13
Suggested readers:	t.b.d.
Version:	1.0
Total number of pages:	69
Keywords:	

Abstract

This document reports on the WP4 activities conducted during the first year of the OFELIA project.

Disclaimer

Imprint

[Project title]	***OpenFlow in Europe – Linking Infrastructure and Applications***
[short title]	***OFELIA***
[Number and title of work package]	***WP4 – Campus Network and emulation wall***
[Document title]	***First year report on planning development testing and operation of individual islands***
[Editor]	***Wouter Tavernier, IBBT***
[Work package leader]	***Didier Colle, IBBT***
[Task leader]	
[PM (estimated)]	***7***
[PM (consumed)]	***7***

List of authors

Organisation/Company	MM	Author
IBBT		**Didier Colle**
IBBT		**Bart Jooris**
IBBT		**Brecht Vermeulen**
IBBT		**Wouter Tavernier**
ETHZ		**Jose Francisco Mingorance-Puga**
ETHZ		**Wolfgang Mühlbauer**
I2CAT		**Leonardo Bergesio**
I2CAT		**Marc Suñé**
TUB		**Marc Körner**
TUB		**Herbert Almus**
UEssex		**Ramanujam Jayakumar**
UEssex		**Nikolaos Efstathiou**
UEssex		**Mayur Channegowda**
UEssex		**Siamak Azodolmolky**
CREATE-NET		**Matteo Gerola**
CREATE-NET		**Roberto Doriguzzi**
CREATE-NET		**Elio Salvadori**
EICT		**Andreas Köpsel**
NEC		**Thomas Dietz**
CNIT		**Giacomo Morabito**
CNIT		**Stefano Salsano**

Table of Contents

List of figures and/or list of tables

Abbreviations

OFELIA	OpenFlow in Europe – Linking Infrastructure and Applications
OF	OpenFlow
ES	End system
CS	Control system
MS	Management switch
NEC	NEC Europe Limited
SW	Software
HW	Hardware
TUB	Technische Universität Berlin
VM	Virtual Machine
WC	Wiring center
FR	Franklinstraße
CIT	Complex and Distributed IT Systems
VeRTIGO	ViRtual TopologIes Generalization in OpenFlow networks

1 Executive summary

The first phase of the project was reached at M7 (March 2011). Since then the individual islands have been up and running with basic functionalities. At that moment, development plans have also been articulated for the next phase (March 2012) for each island.

The document starts in a first section with a brief general discussion. It highlights the basic architecture/concept from which each individual island was developed. It also gives a brief overview of the tests that have been conducted to commission the individual islands.

Each of the remaining sections then discusses a particular island in more detail. First the five islands (IBBT, ETHZ, I2CAT, TUB and UESSEX islands) are discussed which were planned from the very beginning, describing the current status reached so far, reporting on operational issues encountered since the commissioning of the islands in Phase I and finally what the next steps are for further developing the island during Phase II.

Then two islands contributed by EICT and NEC are briefly described. These islands are assumed to be non-production islands, giving the opportunity to test new functionalities under development in the OFELIA project.

Finally, the two islands are discussed that will be contributed by the new partners that joined the OFELIA consortium during the first round of open calls. As these islands are not operational yet, only plans for Phase II are given.

2 Island architecture / concept

The purpose of this section is to give some insights in the generic architectural/conceptual assumptions from which the individual islands have been developed. A first section highlights how each island is built up from three logical networks (experimental, control and management). A second subsection focuses on the control network and its cross-island architecture, addressing scheme and routing. The third subsection discusses some open issues for which operational experience should be obtained before a final decision can be taken. Finally, the fourth section gives a brief overview of the test plans common to all islands that were defined in order to commission the individual islands.

2.1 Overall logical network architecture per island.

As depicted by Figure 1, each island can have up to three logical networks:

1. **Experimental network** (black lines in the figure): this is the network that interconnects the OpenFlow switches and Virtual Machines (VMs) available for experimenting. The experimental network resources are the network resources that are sliced by the control framework.
2. **Control network** (blue lines in the figure): besides the experimental infrastructure, the facility also contains some servers to control and use the facility. Besides the OFELIA specific servers (expedient, opt-in mgr and VT-AM, FlowVisor), also more generic services are provided like an LDAP server for storing the credentials, DNS server, NFS server, etc. The purpose of this control network is to interconnect these servers and the infrastructure under control, plus it gives the users access to their experiments (slices). The left side of the figure shows how the control network may also be connected to other local test-beds or to the Internet. The connection to the Internet serves two purposes: first of all, OpenVPN L2 tunnels (later to be replaced by dedicated circuits) interconnect the different islands to the hub, and secondly, some islands may directly allow downloading software packages through a firewall. It is considered that the control network of each island sits behind a gateway router.
3. **Management network** (red lines in the figure): the purpose of the management network is to allow island managers to manage the infrastructure in their island. Users are not expected (allowed) to use the management network, or send traffic over it. As shown at the right side of the figure, this network will typically be directly reachable from the local offices where the island managers are located (most probably, firewalling will take place to protect the local office network).

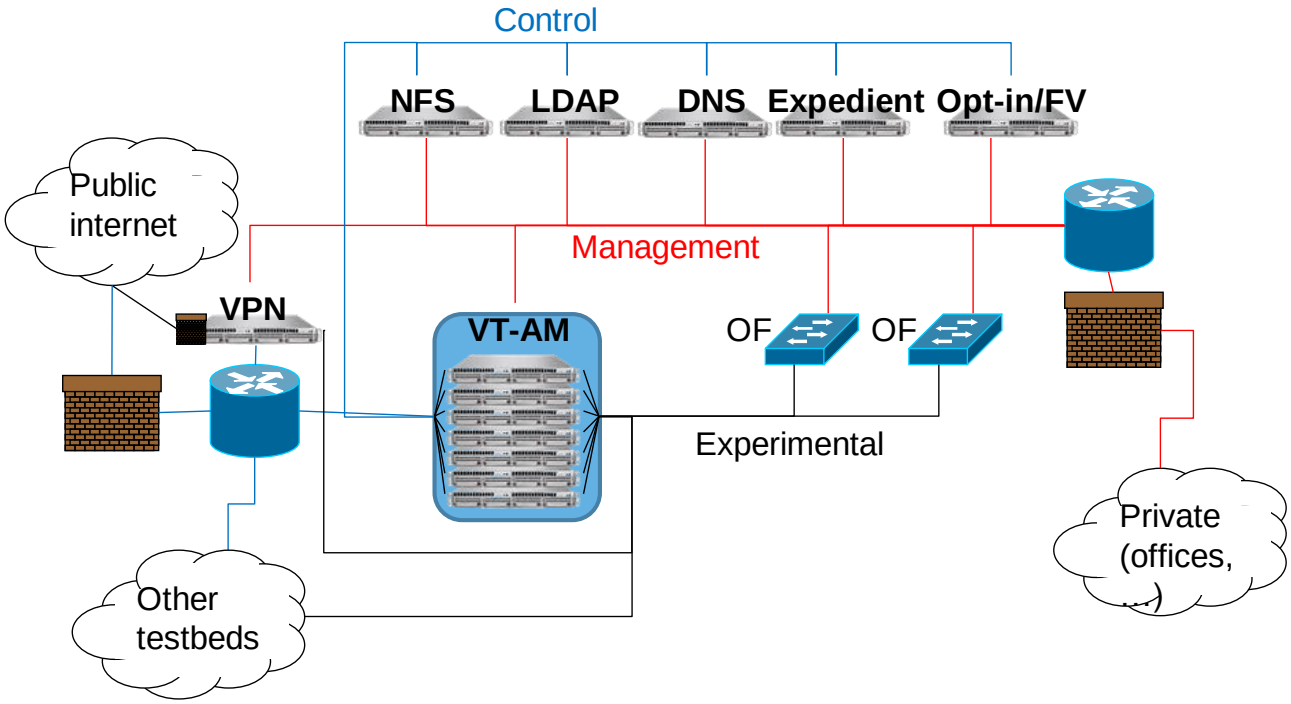

Figure 1: Overall logical network architecture per island

As the control network gets interconnected with the control networks of the other islands (through the L2 OpenVPN tunnels) and to other local test-beds, the control network is the network that requires most coordination in terms of design choices like addressing scheme, routing, etc. These issues will be discussed in more details in the next sections.

Phase I OFELIA does not support inter-island experimental traffic. This will only be supported in later stages of the project.

The management network is so far an optional network: in some islands its function may be fulfilled by the control network (having the advantage of being simpler to implement, but jeopardizing the separation from user traffic).

2.2 The control network: architecture, addressing scheme and routing

As illustrated in Figure 2 and explained in the previous section, the control network of each individual island is located behind a gateway router. This (logical) gateway router may route traffic from/to other local test-beds or via central global subnet traffic from/to other islands. The figure also shows how (external) users may get into the OFELIA facility by establishing L3 OpenVPN tunnels to the hub located in Ghent (IBBT): from there on, they can reach a control network endpoint in any of the islands.

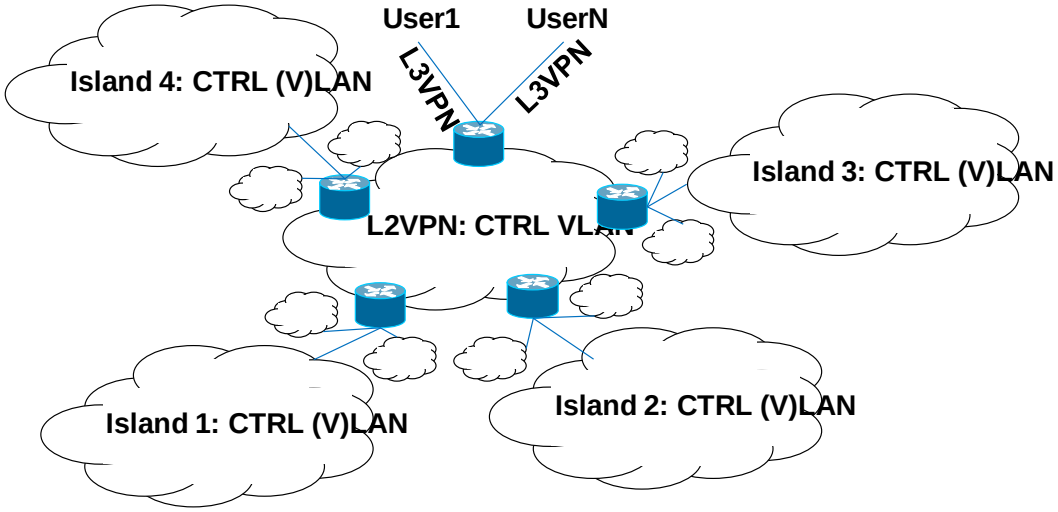

Figure 2: Cross-island control network architecture

To facilitate the routing inside the control network across islands, a common addressing scheme was agreed: this addressing scheme is depicted in Figure 3. Besides the other local test-beds (not under (full) control of OFELIA), all OFELIA control and management network addresses fall in the address range 10.216.0.0/16. The 17th bit divides this address range into a control network address range (10.216.0.0/17) and a management network address range (10.216.128.0/17). The remaining 15 bits are divided into an island (or more generic subnet) ID of 5 bits and a host/interface ID of 10 bits (as CNIT is contributing to two separate islands provisionally, CNIT was assigned two islands IDs: however, this still needs to be confirmed as a trade-off exists between address range usage and operational simplicity). In other words, up to 32 subnets can be assigned to an island (subnet) ID and in each such subnet 1024 host/interface addresses can be assigned. Besides the island ID for each individual island, the central global subnet interconnecting the gateway routers is assigned the address range 10.216.0.0/22, and address range 10.216.124.0/22 (island/subnet ID=31) is allocated for the L3 OpenVPN tunnel endpoints (the VPN tunnels used by external users to get inside the OFELIA facility). The latter range is further subdivided into the range 10.216.124.0/23 for OpenVPN tunnels that rely on user-password credentials authentication, whereas 10.216.126.0/23 is reserved for OpenVPN tunnels that rely on X509 certificates for authentication.

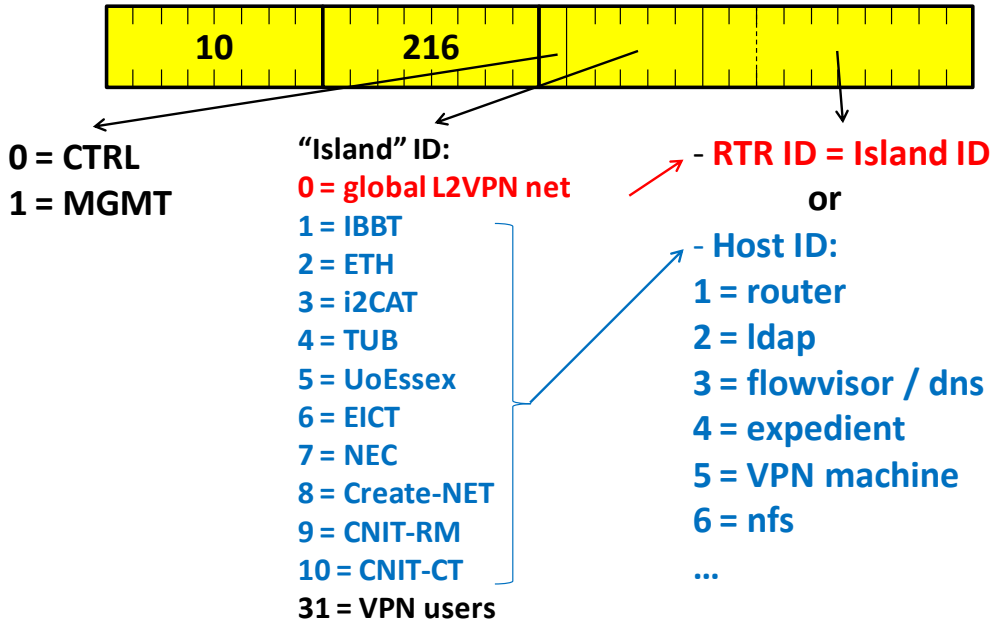

Figure 3: Control network addressing scheme

Given the architecture described in Figure 2 and the addressing scheme in Figure 3, routing should take place as follows:

- All nodes connected to the control network, except the gateway routers: should point to the proper gateway router interface as next hop for all control network addresses (at least 10.216.0.0/17) beyond the subnet they belong to, plus for the address ranges of any other test-bed that should be reachable and is connected to the OFELIA facility. Alternatively, one may also install this next-hop as default router after installing any routes that should go over the management network and/or experimental network.
- Gateway routers: besides the address ranges of any local test-bed connected to the gateway router and any control network subnet directly connected to the gateway router, it should point to the proper opposite router interface as next hop for all address ranges reachable through that opposite gateway router. Thus for example, the next hop for the control network subnet of island with ID <x> (address range 10.216.<x>*4.0/22) is 10.216.0.<x>.

2.3 Open issues

In this section, we briefly discuss a couple of open issues that have been debated extensively within the OFELIA consortium, but for which no consensus was reached and thus decision is postponed until sufficient operational experience exists to take a final decision.

- **Slicing mechanism**: several options to perform the slicing (or identifying to what slice a packet/frame belongs) exists. The main options considered by the OFELIA consortium are slicing based on VLAN-ID, or based on MAC address. Whereas the MAC-based slicing has the advantage that it enables experiments involving the VLAN-ID (e.g., (G)ELS or GMPLS-based Ethernet Label Switching experiments), it suffers from the fact that in the worst case the number of TCAM-entries used in a slices scales quadratically with the number of MAC-addresses in that slice. On the other hand, VLAN-based slicing only multiplies the number of required TCAM entries by the number of VLAN-entries in a slice (which would often equal one, and a small number in exceptional cases). Another consideration to take into account is that VLANs are already used in some places to separate traffic/experiments/…
- **RSI vs. VSI mode**: the NEC switches used in the OFELIA project can provide OpenFlow switch instance(s) in two modes. In the first mode, the Real Switch Instance (RSI) mode, only a single OpenFlow switch instance can be configured on the box by including some specific switch ports in the OpenFlow switch instance and slicing should be performed by an external FlowVisor. In the

second mode, the Virtual Switch Instance (VSI) mode, on each switch, up to 16 OpenFlow Virtual Switch Instances (VSIs) can be configured: a VSI is configured by specifying the VLAN-ID of the VSI. Whereas the RSI mode has the advantage of not imposing restrictions on the slicing mechanisms to use and not being limited to only 16 OpenFlow switch instances, the VSI mode may have the advantage that no external FlowVisor would be needed to perform the slicing (as in this case the slicing is done inside the box), potentially resulting in better performance.

- **Implementation of dedicated management network**: in the description above, it was already mentioned that implementing a dedicated management network is not mandatory when its functions are fulfilled by the control network. Not implementing another dedicated management network probably reduces complexity, but may be harder to shield the management services from (misbehaving) users.

2.4 Island commissioning: common testing procedures

At the end of Phase I, a common testing plan was put in place for the commissioning of each individual island before the island was considered to be in a production state. All islands have passed these tests successfully as part of milestone MS46.

First of all, a basic configuration test suite consisting of tests to validate connectivity with the management network, and tests to validate the connectivity and traffic separation within basic OpenFlow scenarios, consisting of one or multiple slices, were defined.

Secondly, a test suite for testing the configuration process within a single OFELIA island was detailed. At one side, this consists of tests dealing with aspects regarding Island Manager configuration, while at the other side this consists of test dealing with User use case configuration aspects such as the configuration of the OpenFlow controller and management of (virtual) resources.

The details of the test procedures as specified in milestone MS44 are left out of this document. Nevertheless, below the templates used by each Island Manager to conduct the commissioning tests are given.

Table 1: Basic configuration test suite template

Test Item	Test Name	Short Description	Result	Remarks
1.1.1	Ping Test	Install OF switch and verify connectivity to the management network.	☐ Success ☐ Failure	
1.1.2	SSH Test	Configure and test SSH connection to manage the switch from a remote location.	☐ Success ☐ Failure	
1.1.3	Control System Connection Test	Configure and test connection to the OFELIA control system.	☐ Success ☐ Failure	
1.2	Single OpenFlow Scenario	Verify basic OpenFlow scenario	☐ Success ☐ Failure	
1.3	Single Slice Testing	Setup a single slice and verify connectivity and traffic separation	☐ Success ☐ Failure	
1.4	Multiple Slice Testing	Setup multiple slices and verify connectivity and traffic separation between slices	☐ Success ☐ Failure	

Table 2: Island Manager (IM) configuration test suite template

Test Item	Action	Short Description	Result	Remarks

2.2.1	Virtualization AM configuration	The IM adds servers to the Virtualization AM.	☐ Success ☐ Failure	
2.2.2	FlowVisor	FlowVisor configurations and instantiation.	☐ Success ☐ Failure	
2.2.3	Opt-in Manager configuration	The IM sets a Clearinghouse user and a FlowVisor	☐ Success ☐ Failure	
2.2.4	OFELIA CF configuration: Virtualization AM	The IM adds the Virtualization AM in the Expedient.	☐ Success ☐ Failure	
2.2.5	OFELIA CF configuration: Opt-in Manager	The IM adds the Opt-in Manager in the Expedient.	☐ Success ☐ Failure	

Table 3: User use case configuration test suite template

Test Item	Action	Short Description	Result	Remarks
2.3.1.a	Create and configure project	The user access OFELIA CF (Expedient). Creates a project and adds AMs.	☐ Success ☐ Failure	
2.3.1.b	Create and configure slice	The user access OFELIA CF (Expedient). Creates a slice and adds AMs.	☐ Success ☐ Failure	
2.3.2	Set slice controller	OpenFlow controller is added for the slice	☐ Success ☐ Failure	
2.3.3	Manage slice resources	Visualize available resources	☐ Success ☐ Failure	
2.3.4	Create a VM	A VM is created in one of the available servers. This should be done twice for the test.	☐ Success ☐ Failure	
2.3.7	Allocate OpenFlow resources	The proper flow (connections) is selected in order to conduct the test.	☐ Success ☐ Failure	
2.3.8	Flowspace request approval	The IM adds a rule to accept requested flowspace.	☐ Success ☐ Failure	
2.3.9	Ping Test	Conduct the ping between VMs	☐ Success ☐ Failure	

3 IBBT island

3.1 IBBT island: current status

An overview of the IBBT island as deployed in Phase I is given in Figure 4. This is almost exactly as planned in MS4.2, except that the 10G link between the NEC hub switch and the iLab.t Virtual Wall Force10 switch is not deployed yet (as it is not needed as long as no cross-island experiments are needed).

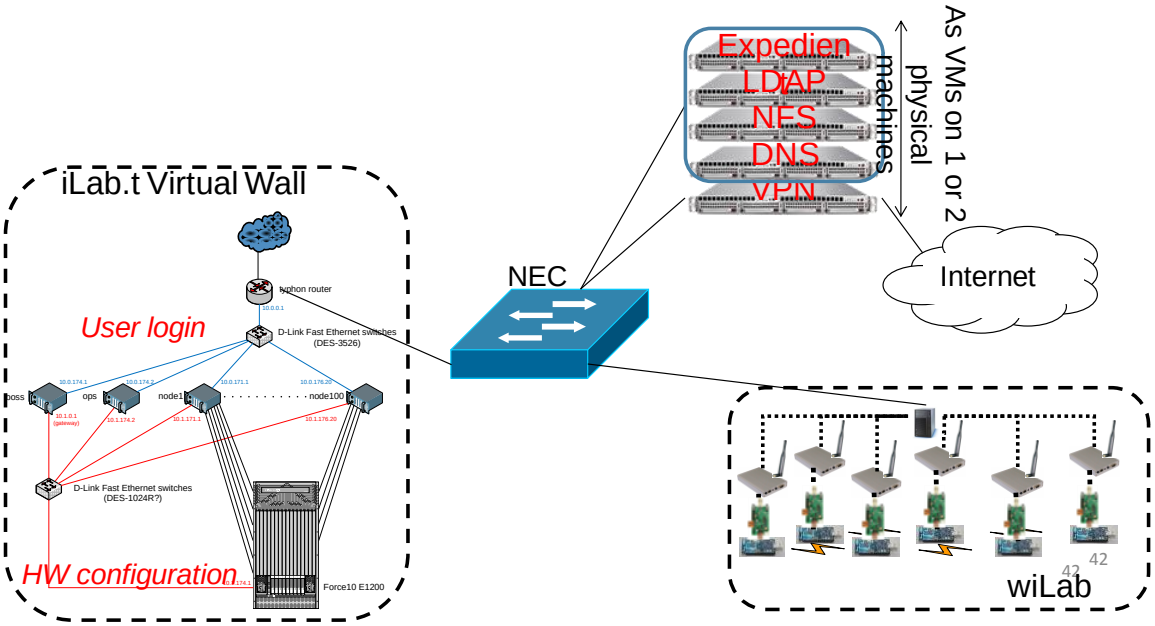

Figure 4: Overview IBBT island Phase I

In summary, the figure shows that the IBBT island after Phase I consists of 3 parts:

1. iLab.t Virtual Wall
2. w-iLab.t Wireless Test-bed
3. OFELIA hub infrastructure

The OFELIA hub infrastructure is the part that has been added during Phase I, compared to the initial state of the island, that had already the iLab.t Virtual Wall and w-iLab.t Wireless Testbed that is shared amongst several projects.

The following subsections will give more details about each of these three parts.

3.1.1 Hub Infrastructure

In the current state (Phase I) of the IBBT island, the hub infrastructure consists of one NEC IP8800/S3640-48T2xW-LW switch that serves as hub switch for the whole OFELIA facility and two servers. One server is fully dedicated to handle the OpenVPN traffic. The other server runs several virtual machines, each acting as a specific hub infrastructure server: KVM is used for the virtualization of the physical server. In case of capacity shortage, server capacity will be extended by moving some of these virtual machines to another physical server.

There are four servers deployed:

1. An LDAP server containing user information including password and (net)group information.
2. A DNS server that forwards requests to either the OFELIA DNS server deployed in the EICT island or the IBBT test lab DNS server for resolving for example names to access nodes on the iLab.t Virtual Wall.
3. An expedient server

4. An NFS server allowing users to temporary store data beyond the lifetime of their experiments. The storage itself is actually provided by an IBBT internal storage server. Currently, OFELIA has an NFS share on that IBBT internal storage server of 500GB: when needed, this storage capacity can be extended. Later on, storage will be migrated to an i-SCSI server over a dedicated network.

Besides these four production servers, temporary additional virtual machines can be deployed for development, testing, etc., purposes.

Figure 5 shows the detailed logical network configuration of the hub infrastructure. At the top in the middle, the NEC hub switch can be seen, at the left side the OpenVPN server and at the right side the hub infrastructure servers (VMs on single physical server). At the bottom, the iLab.t Virtual Wall and w-iLab.t Wireless Testbed are shown. In the middle, the IBBT typhoon firewall is shown: this firewall sits in between several test networks and is the exit/entry point from/for these test networks from the local IBBT offices and to the Internet (although access to the Internet has been restricted in such a way that traffic should go through a proxy server). The OFELIA control network is just another network attached to this firewall.

Currently 4 of the 48 GbE ports are in use on the NEC IP8800/S3640-48T2xW-LW hub switch (in Phase II, 10 NetFGPAs will also connect to this switch and thus require 40 GbE ports).

1. **Port 0/48**: this port with address 10.10.28.10 (on VLAN 110 in access mode) is logically directly attached to the test network 10.10.0.0/17 in which the island manager has some test machines from which the switch was initially configured. Except traffic for this subnet, no traffic is ever routed via this interface, and thus this port is kept as emergency back-door to configure the switch (as long as we do not have to give it up for other purposes).
2. **Port 0/47**: this port with address 10.216.7.253 (on VLAN 101 in access mode) is the link to the local IBBT networks (iLab.t Virtual Wall, w-iLab.t Wireless Testbed, offices, etc) and thus is connected to the IBBT typhoon firewall that has address 10.216.7.254. The subnet 10.216.7.240/28 allocated to this connection was chosen sufficiently large in order to easily accommodate additional machines for example for troubleshooting purposes.
3. **Port 0/46**: is the port directly connected to the physical server running the virtual machines for the hub infrastructure servers. This port operates in trunk mode and currently accommodates two VLANs. VLAN4094 is the IBBT OFELIA management network (with address range 10.216.132.0/22) shown as red lines in the figure. VLAN4092 is the IBBT OFELIA control network (with address range 10.216.4.0/22) shown as blue lines in the figure. At the opposite side (in the server), both VLANs are extended and distributed to the different virtual machines by means of kernel bridges. The management network is also logically terminated on address 10.216.132.125 by the host operating system to allow management of the server (starting/stopping virtual machines, etc).
4. **Port 0/45**: is the port directly connected to the OpenVPN server. Also this port operates in trunk mode and currently accommodates two VLANs. VLAN4094 is the IBBT OFELIA management network (with address range 10.216.132.0/22) shown as red lines in the figure: this is the same VLAN as on port 0/46. As nothing else than the OpenVPN server needs to be managed via this port, this VLAN is simply terminated on address 10.216.132.5 at the opposite side. VLAN4091 (with address range 10.216.7.224/22 and terminated by the hub switch on address 10.216.7.238) transports the traffic from/to the end users who connect to the OFELIA facility via L3 OpenVPN tunnels and the control traffic to/from the control networks in the other islands: at the opposite side the kernel routing functionality terminates this VLAN on address 10.216.7.237. Besides these VLANs, VLAN4090 (currently not configured on the hub switch) is bridged in the OpenVPN server to/from the tap interfaces belonging to the OpenVPN daemons handling experimental traffic.

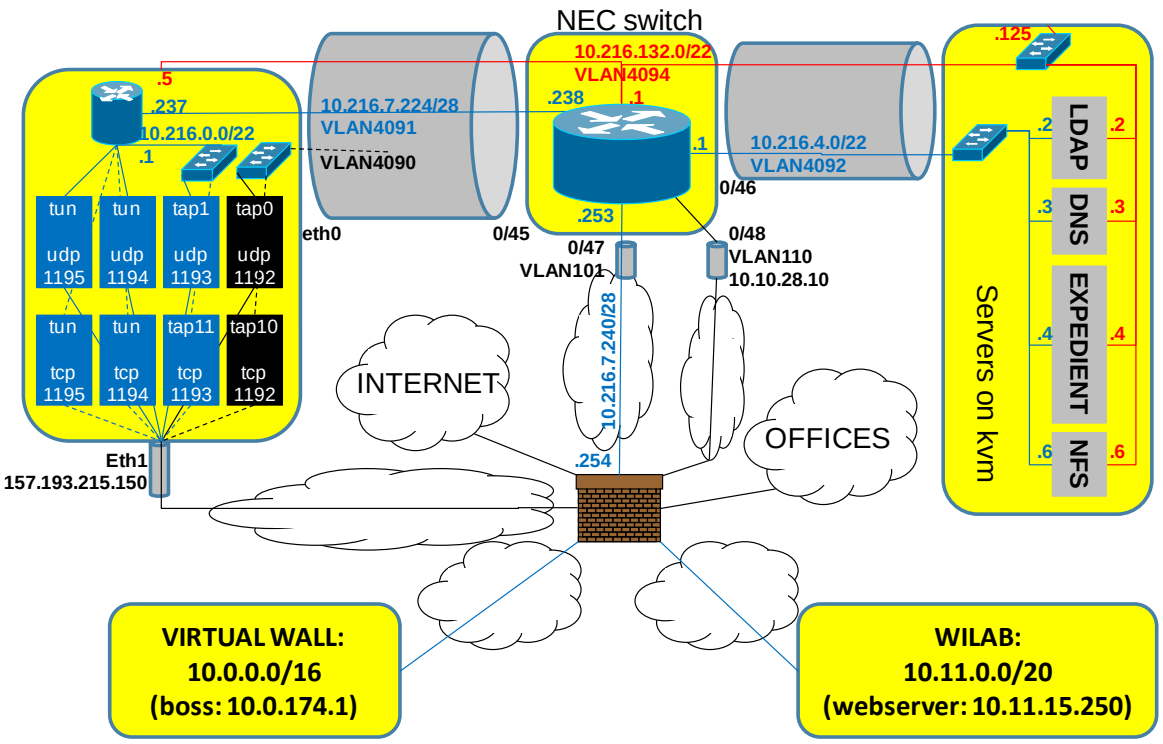

Figure 5: Overall detailed configuration of IBBT island in Phase I

The configuration of the hub switch as described above can also be derived from the routing table as presented in Figure 6.

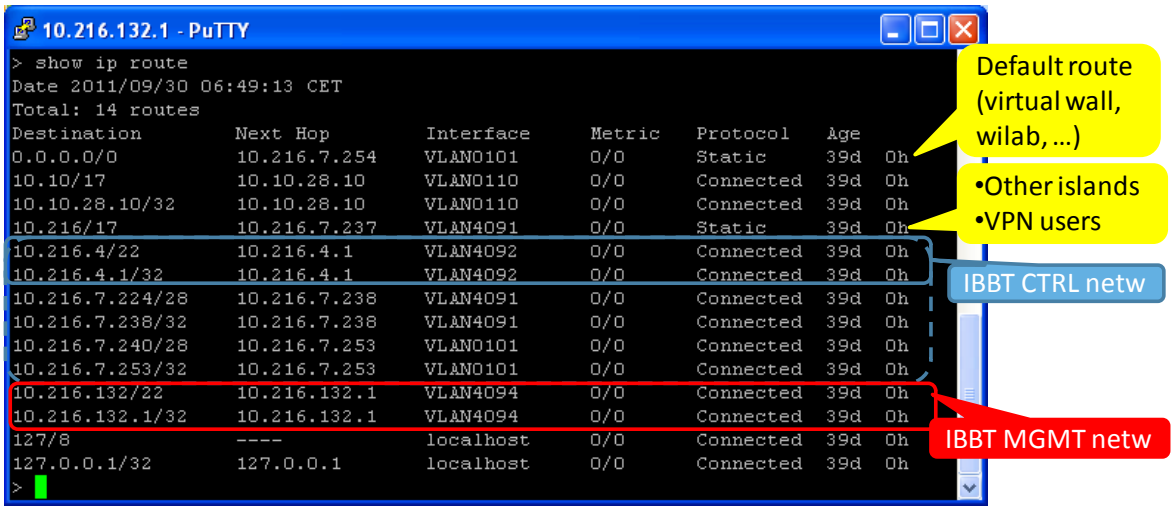

Figure 6: Configuration of Hub Switch

As illustrated in Figure 5, OpenVPN traffic comes/goes through the IBBT typhoon firewall to/from the eth1 port on the OpenVPN server with address 157.193.215.150 which is a publicly routable address. As this figure shows, the server runs two times 4 OpenVPN daemons:

1. **UDP and TCP Ports 1194 (= default OpenVPN port)**: this daemon accepts L3 OpenVPN connections, allowing external users to get into the OFELIA facility. Authentication by this daemon is done by checking the user/password credentials against the LDAP server.
2. **UDP and TCP Ports 1195**: this daemon also accepts L3 OpenVPN connections to allow external users getting into the OFELIA facility, but authentication is done based on X509 certificates.
3. **UDP and TCP Ports 1193**: this daemon accepts L2 OpenVPN connections from other islands and serves the interconnection of the OFELIA control networks. Client-to-client traffic is allowed by this

daemon. The TAP interfaces belonging to these daemons are tap1 and tap11 which are bridged to the kernel routing functionality on address 10.216.0.1, forming a L2 network with subnet 10.216.0.0/22.

4. **UDP and TCP Ports 1192**: this daemon also accepts L2 OpenVPN connections from other islands and serves the transport of experimental traffic (however, inter-island experimental traffic is not supposed to be supported already). Also in this case client-to-client traffic is allowed and the accompanying TAP interfaces are tap0 and tap10.

Although transport over UDP should provide better performance, TCP daemons were introduced as alternative option as the unreliable transport over UDP may rarely result in breaking the SSL connection used for the encryption of the tunnels.

As eth1 is a publicly reachable interface and the only means to get inside the OFELIA facility is through OpenVPN connections, firewall rules have been set to drop all traffic except traffic on the above mentioned UDP and TCP ports locally received/generated on this port. To avoid that these publicly routed packets are routed to the central hub switch (due to default routing), these firewall rules also mark these packets such that a higher priority routing table can be used to send them out of the eth1 interface facing the public Internet.

3.1.2 iLab.t Virtual Wall

This section describes the IBBT iLab.t Virtual Wall. The first subsection gives a brief overview of changes and/or customization specific to the OFELIA project that happened during Phase I. The second subsection describes the iLab.t Virtual Wall infrastructure in full detail.

3.1.2.1 Phase I modifications

An OpenFlow-enabled image for the IBBT iLab.t Virtual Wall has been created. This image features at one side the OpenVSwitch version 1.1.0 and at the other side the NOX OpenFlow controller version 0.9.0 (Zaku) both supporting the OpenFlow version 1.0. The image also contains a script for automatically deriving which experimental ports are supposed to be added to the datapath when starting the OpenVSwitch and which port is the port connecting (possibly via a LAN node) to the OpenFlow controller. Finally this OFELIA image features the kernel patch to translate EtherType 0x8100 into 0x9100 when sending out an Ethernet frame and vice versa when receiving an Ethernet frame: in this way, experiments involving VLAN ID manipulation are possible although the Force10 switch is separating experiments by means of configuring port based VLANs. Of course, this implies that EtherType 0x9100 has become unusable: as this EtherType is far less used, it is expected that this will very rarely or even never cause a problem.

An expedient plug-in has been developed to enable defining and swapping in and out experiments on the IBBT iLab.t Virtual Wall. For this purpose, one of the Emulab GUI applets was integrated into expedient: this applet allows the user drawing a topology and accordingly generating a NS-script that can be handled by the Emulab software on the Virtual Wall. The plug-in also allows afterwards adapting the NS-script, which is useful for example for starting the OpenVSwitch or NOX controller on particular nodes: we refer the user to the user manual pages for more details how this can be done.

3.1.2.2 Detailed description

Technical description

The iLab.t Virtual Wall is shown in Figure 7. This infrastructure consist of 100 general purpose machines/servers, each featuring dual CPU dual core processors at 2.0 GHz, 4 GB RAM, 4 disks of 80 GB, IPMI with two network interfaces and 6x 1GbE network interfaces (PCI-E) that are available for experimental use. For 40% of the nodes, 2 experimental NICs are unconnected; all other experimental NICs (60x6+40x4=500 in total) are connected to a Force10 E1200 switch that features 576x 1GbE, 8x 10 GbE ports and a non-blocking backplane at 1.68 TBps. Such a setup allows flexibly interconnecting a set out of the available general purpose machines/servers in many different topologies (ring, mesh, star, etc). 20% of the servers also features a graphics card and attached display, needed for video related experiments.

The purpose of the iLab.t Virtual Wall is two-fold. First of all, the purpose is providing a large scale environment for enabling experimental network research. Secondly, the intention is providing an infrastructure that is as generic as possible, in order to avoid as much as possible restrictions on possible

experiments, in order to increase the usage of the infrastructure as much as possible which is critical in motivating such a huge investment.

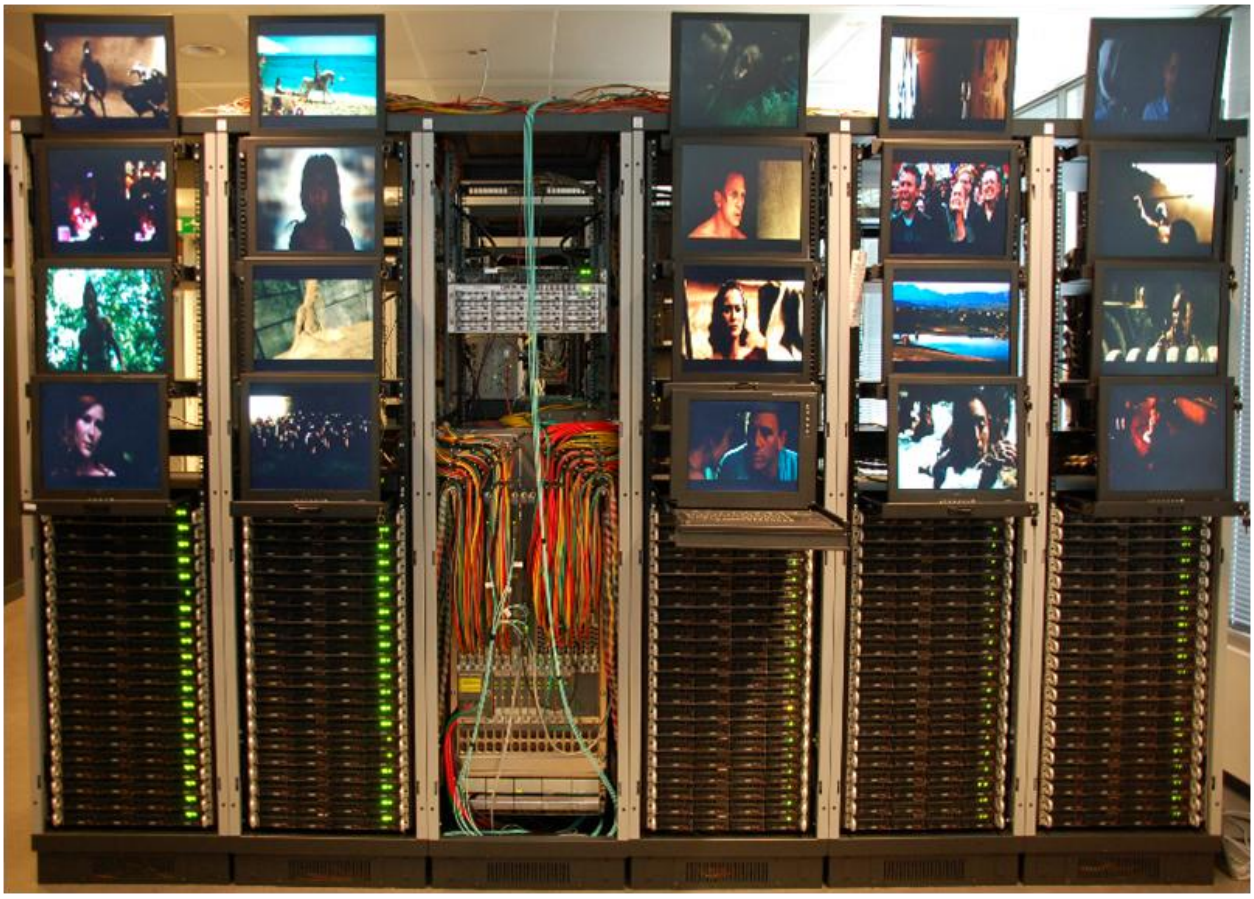

Figure 7: iLab.t Virtual Wall

The network architecture of the iLab.t Virtual Wall is depicted in Figure 8. As mentioned above, the core of the iLab.t Virtual Wall is the Force10 switch connecting each of the 100 servers through 4 or 6 1GbE links. Besides that there are two other machines: boss and ops. Boss is the machine where the Emulab software (developed by the University of Utah) runs and that is in charge of controlling the whole infrastructure and that provides a web and xml-rpc interfaces to the experimenters using the infrastructure. Ops provides a file share where data can be stored that can be mounted automatically on the experimental nodes during executing experiments.

Users log in on the Emulab web interface to define and control their experiments. Defining experiments basically boils down to defining a topology (nodes and links) and specifying which OS images that needs to be booted on which nodes and what scripts should be automatically started once the nodes are booted. It is also possible to define nodes as virtual rather than physical nodes, allowing larger scale experiments with the same amount of physical resources: experience has shown that often up to 20 virtual nodes can be mapped onto a physical node, although this is strongly dependent on the actual experiments and the actual resources the virtual nodes consume. Besides a GUI, experiments can be specified through an NS-like scripting language. The same interface can be used to swap-in and swap-out experiments.

At swap-in the Emulab software configures the right VLANs on the Force10 switch, in order to keep the experimental traffic of that particular experiment inside the experiment in order to avoid that other experiments are affected by that traffic. Port based VLANs are used for physical nodes, avoiding the need to tag traffic in those nodes. The Emulab software also reboots the nodes in the experiment through one of the network interfaces on the IPMI board. A dedicated configuration network (in the address range 10.1.0.0/16) of D-Link DES-1024R Fast Ethernet switches serves the purpose of configuring VLANs on the Force10 switch and rebooting the nodes. The Emulab software makes use of Frisbee to efficiently distribute the OS images to be loaded on the nodes through the wall.test control network (in the address range 10.0.0.0/16) that consists of D-Link DES-3526 Fast Ethernet switches. Although experiments are also allowed to assemble

their own customized OS images to be loaded on the nodes in their experiments, it is often recommended to make use of the default OS images and to (automatically) install the specific packages needed for the particular experiment each time the experiment is swapped in. The wall.test control network is also the network through which experiments can get access to the boss and ops nodes and to the nodes in their experiments (e.g., through SSH). The wall.test control network is a Fast Ethernet network, thus not really a high-speed network offering no QoS guarantees: therefore, experimental traffic on this network should be avoided as much as possible.

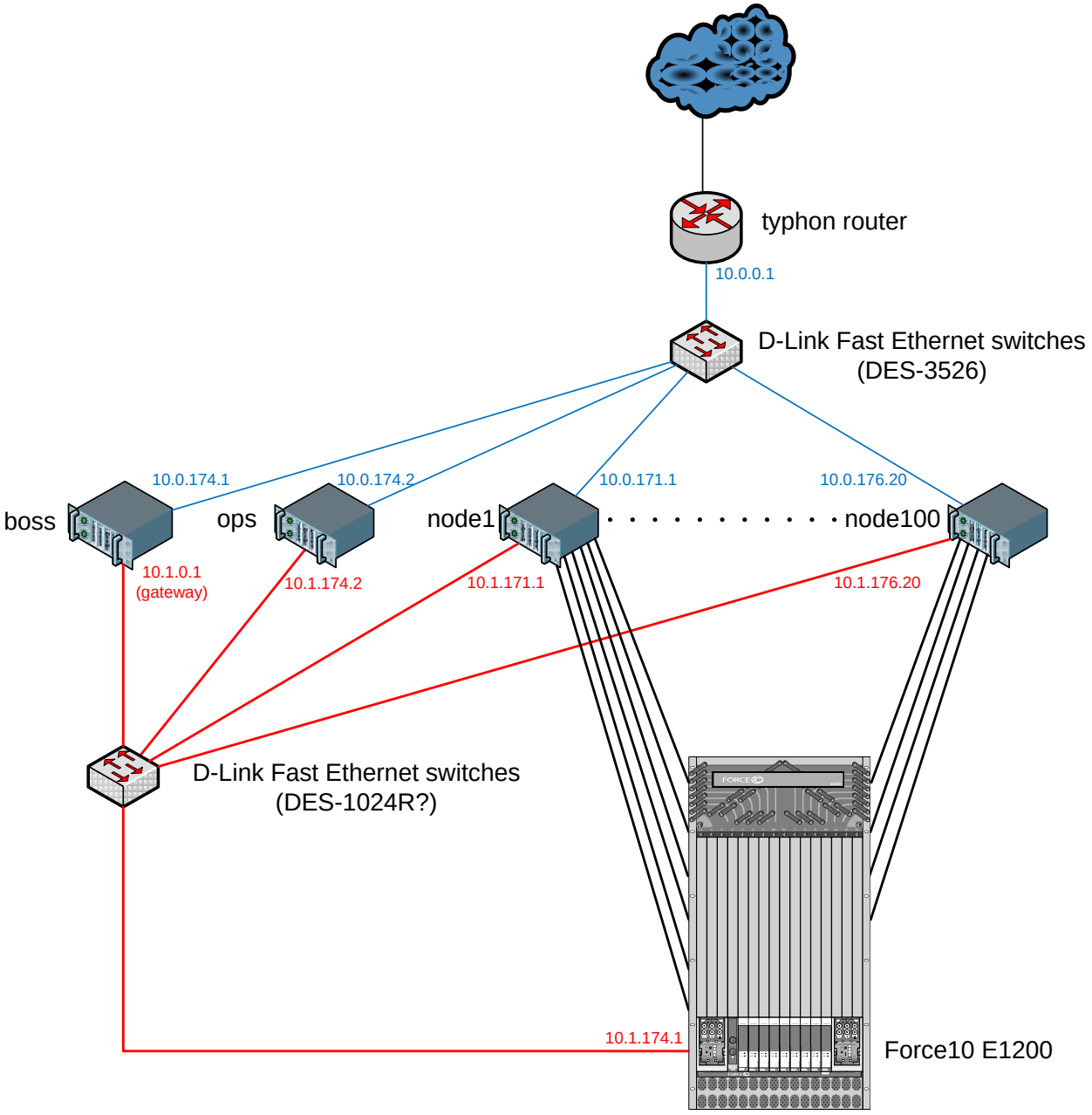

Figure 8: iLab.t Virtual Wall network architecture

What is offered to the OFELIA project?

OFELIA and its users, will of course be allowed to register on the iLab.t Virtual Wall and define and control experiments as regular users through the Emulab front-end, despite the fact that this mode of operation is not the one envisaged by the OFELIA project, that would have its own (expedient based) front-end.

Besides that, OFELIA is offered following two things:

- Access through the xml-rpc based APIs offered by the Emulab software deployed on the iLab.t Virtual Wall. Since MS4.1, an upgrade to the latest Emulab software took place. This latest version

of the Emulab software features a ProtoGENI interface, being the Emulab incarnation of the SFA architecture (Aggregate Manager (AM) interface). However, so far attempts to get this ProtoGENI interface operational were unsuccessful. Therefore, currently the Virtual Wall expedient plug-in relies on the proprietary Emulab xml-rpc and further investigation to migrate to the ProtoGENI interface will be further investigated during Phase II of the OFELIA project.

- A physical port to connect the OFELIA infrastructure to the wall.test control network.
- A 10G port on the Force10 switch in the virtual wall to enable experiments beyond the iLab.t Virtual Wall boundaries. A typical example would be deploying OpenFlow controllers on the iLab.t Virtual Wall controlling (slices on) OpenFlow switches in other OFELIA islands or software versions deployed on the w-iLab.t wireless/sensor network testbed.

Constraints to be considered by the OFELIA project

Given the current deployment and policies of the iLab.t Virtual Wall, following constraints have to be taken into account:

- So far, after upgrading of the Emulab software to a ProtoGENI capable version enabling the ProtoGENI interface was not successful. Therefore, during OFELIA Phase II the ProtoGENI interface will be further investigated. And therefore, the proprietary xml-rpc is currently used as workaround solution as long as the technical problems with the ProtoGENI interface have not been fixed.
- VLANs are used to isolate experiments from each other. These VLANs will need to be extended on the 10G port that will connect the Force10 switch to the OFELIA infrastructure, for experiments that reach beyond the iLab.t Virtual Wall boundaries.
    - The Emulab software on the iLab.t Virtual Wall is responsible for allocating the VLAN-IDs to experiments and this allocation may take place as late as the swap-in of the experiment. The (expedient based) OFELIA control software should thus be capable of dealing with this fact.
    - It seems that the current ProtoGENI interface is not communicating this VLAN-ID allocation, even not through the manifest-rspecs. OFELIA shall thus need to extend the ProtoGENI interface with this feature.
- The ProtoGENI interface assumes that links from an experimental node to an external node is realized through a GRE tunnel coming in through the wall.test control network. Further investigation is required to find out how this can be matched with the configuration of the VLANs on the 10G port connecting the OFELIA infrastructure to the Force10 switch in the iLab.t Virtual Wall.
- As VLANs configured on the Force10 switch isolate experiments from each other, experiments making use of VLAN tagging typically run into problems. For this purpose, recently a kernel patch was made that translates at an output port the EtherType into an EtherType that can pass the Force10 switch transparently and vice verso at the input port.
- Defining experiments on the iLab.t Virtual Wall also involves defining a topology. On a particular node, which network interface (which /dev/ethX) is terminating what link of that topology is only decided by the Emulab software at the time of swapping in an experiment. Solutions exists to retrieve this information while an experiment is running (this requires an IP address was assigned to the link: assigning dummy IP addresses if no IP address is needed works perfect).
- The iLab.t Virtual Wall makes use of the 10.0.0.0/16 address range for its control network wall.test and of the 10.1.0.0/16 address range for its configuration network, preventing OFELIA of using this address range for other purposes.
- Assignment of the IP address to the control interface of each node in the wall.test control network is under control of the Emulab software and this assignment may occur as late as swap-in of the experiment: the (expedient based) OFELIA control software should be designed in such way that this fact is taken into account.

- The other interfaces (experimental interfaces on the nodes connected to the Force10 switch) need to be assigned addresses in a /24 subnet in the range 172.16.0.0/24 and 172.31.0.0/24. These IP addresses can be specified at the time of defining the experiment.
- As the iLab.t Virtual Wall is a (time-)shared infrastructure, OFELIA and its users will be subject to the usage policies that will be adopted in the reservation system under development. Until this reservation system will become operational and integrated in OFELIA (which is planned for the second phase of the OFELIA project), no hard/automated enforcement of reservations are possible.

3.1.3 w-iLab.t Wireless Testbed

This section described the IBBT w-iLab.t Wireless Testbed. The first subsection gives a brief overview of changes and/or customization specific to the OFELIA project that happened during Phase I. The second subsection is a detailed up to date description of the w-iLab.t infrastructure and what it is providing to the OFELIA project.

3.1.3.1 Phase I modifications

OpenFlow 1.0 enabled OpenWRT image

To serve the needs for OFELIA, an OpenWRT based image has been created that includes the OpenFlow 1.0 reference implementation (currently, OpenVSwitch is not supported). This image had to be created from a bare OpenWRT distribution, as the publicly available OpenFlow capable OpenWRT images that run on an ALIX embedded PC includes older version of OpenFlow.

To facilitate running an OpenFlow experiment, a script was created that allows configuring the wireless interfaces and possibly starting the OpenFlow switch. In case of an endpoint, the script is executed as follows:

```
/etc/wireless_script_OpenFlow.sh host wlan<0|1> <IP address> <channel #>
```

The host argument refers to the fact that only a WiFi interface is configured and that not an OpenFlow switch is started. As there are two wireless interfaces available, one can choose between wlan0 and wlan1. The IP address is the address assigned to the specified wireless interface. The channel number is the channel (frequency) that the wireless interface is tuned to. Behind the scene, iwconfig and ifconfig commands are called. The user is of course free to call the iwconfig and ifconfig commands himself and tune more of the configurable parameters.

To start an OpenFlow switch, the script is executed as follows:

```
/etc/wireless_script_OpenFlow.sh of <OF-CTRL address> wlan0 <channel #> wlan1
<channel #>
```

The datapath corresponding to the OF switch created is *unix:/var/run/dp0* and thus information can be retrieved by command calls like:

```
dpctl dump-ports unix:/var/run/dp0
dpctl dump-flows unix:/var/run/dp0
```

Python web scraping API

To support OFELIA, also a Python web scraping library has been built. The purpose of this library is to provide an API that automatically interacts with the w-iLab web server and automatically fills in and submits the web forms to schedule experiments on the w-iLab.t wireless testbed. This API will serve as core for a w-iLab.t specific expedient plug-in that is planned in the near future.

OMF migration plan

It is planned to migrate the control framework of the w-iLab.t wireless test facility to an OMF based control framework. Testing and customization are currently ongoing: at this moment, the migration is planned for the second half of 2011. Once the migration took place, the w-iLab.t specific components in the OFELIA control framework will need to be modified in such a way that it can interact with this OMF framework that will control the w-iLab.t test facility. The use of the SFA-based interface provided by the OMF framework will be investigated.

New address range

As described in MS4.1, the w-iLab.t wireless test network made internally use of a private address range, and thus access from the outside world happened through a dynamic NAT service. Since MS4.1, a new address range 10.11.0.0/20 has been allocated so that passing through the NAT will not be required anymore. Until now, only the w-iLab central web server and wilabfs file share are using their new address: the migration of all individual iNodes is planned for the near future.

3.1.3.2 Detailed description

Technical description

The test bed deployed consists out of 200 nodes (according to the status page of the w-iLab.t testbed) spread over three floors of a 15x91m office building. The following figures illustrate how the nodes are spread across the three floors according to the floor plans. Colors indicate different zones in which experiments can be scheduled.

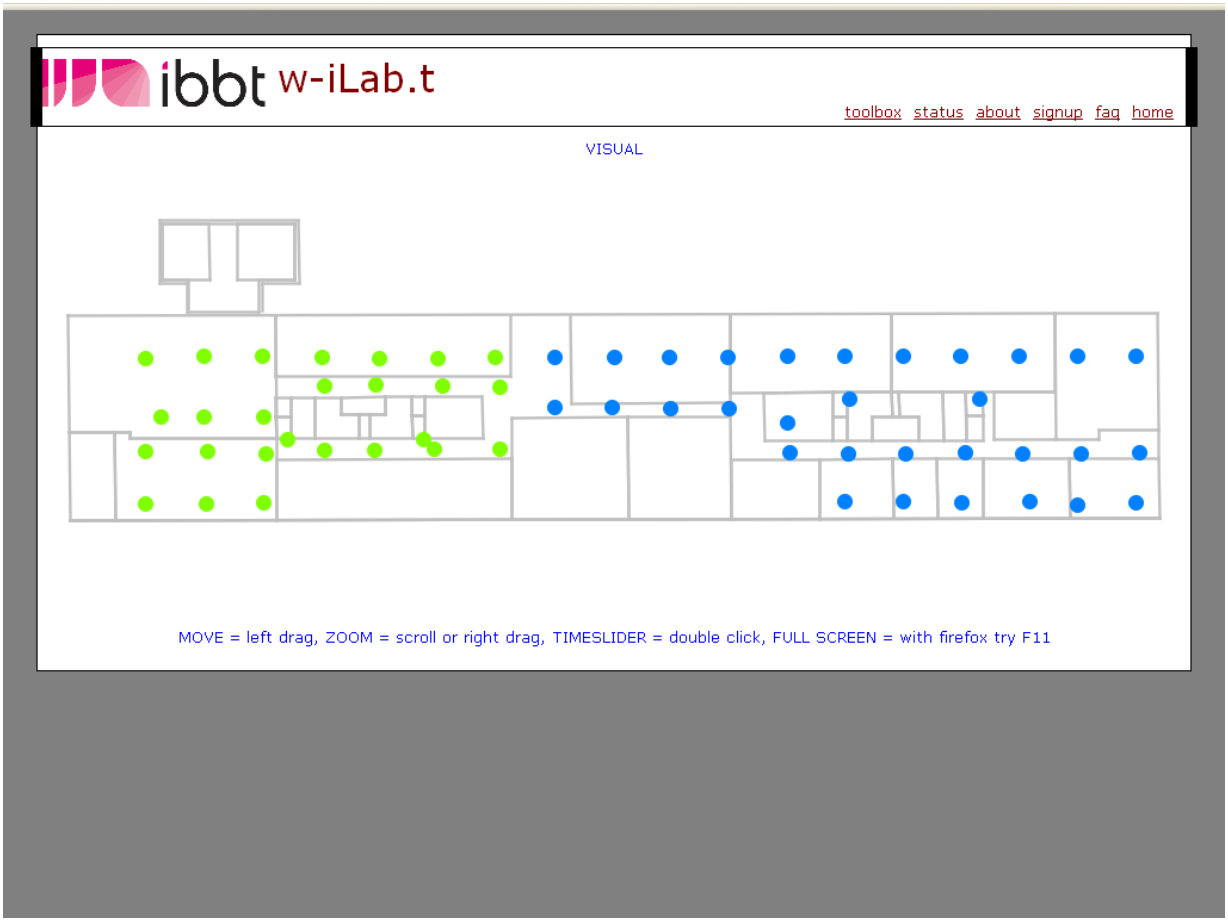

Figure 9: w-iLab.t on 3rd floor.

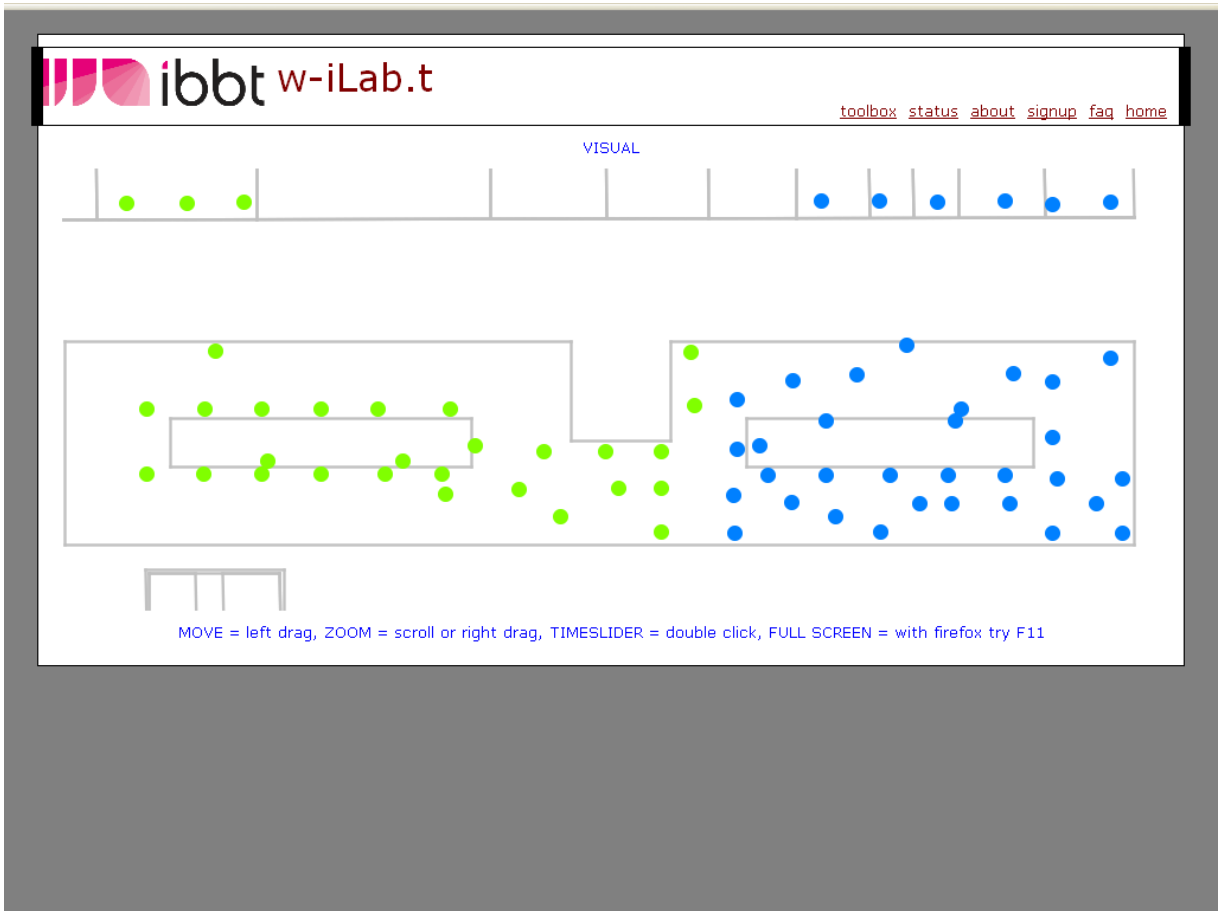

Figure 10: w-iLab.t on 2nd floor

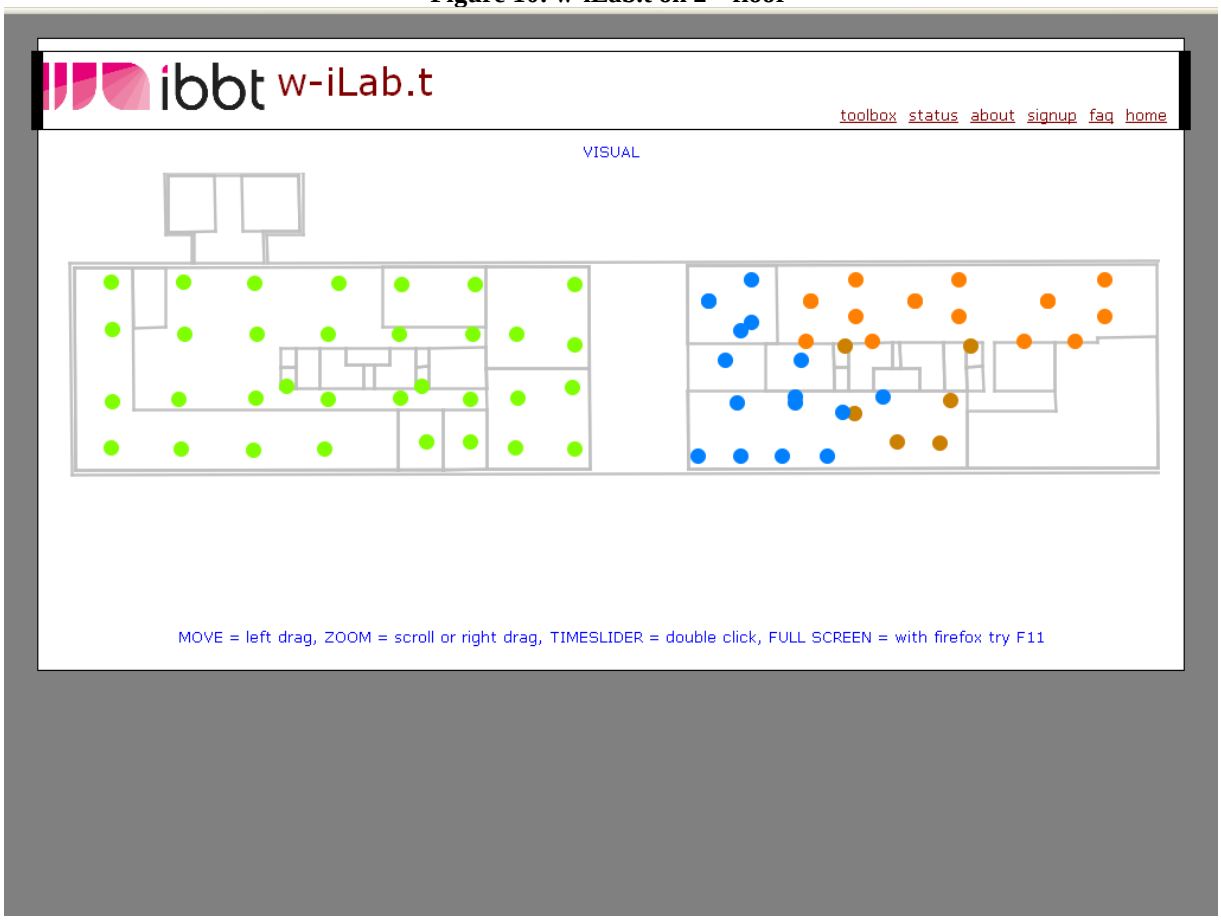

Figure 11: w-iLab.t on 1st floor.

Each node consists of an embedded PC called iNodes (to which sensor modules called motes are attached: these motes are not assumed to be useful in the context of OFELIA). The iNodes are Alix 3C3 (see also http://pcengines.ch/alix3c3.htm) devices running a Voyage Linux distribution (currently Version: 0.6 (Build Date 20090217)), featuring a 500 MHz AMD Geode LX800 CPU, 256 MB DDR DRAM, 1 GB CF storage and equipped with Ethernet, USB, serial, vga, audio I/O interfaces and two 802.11abg wireless network interfaces (compex wlm54sag23) connected to two 5 dBi dual band antenna. For reducing radiation levels on the 3rd floor 10 dB and on the 2nd floor 20 dB attenuators have been installed on all transmitters.

The w-iLab.t network architecture is illustrated in Figure 12. The w-iLab central server provides a web-based control interface (www.wilab.test) through which researchers can define and schedule their experiments. Behind this server is a couple of HP2600 series switches that physically connect the individual iNode LAN interface to this server. The main purpose of this network is the execution/life-cycle management of the experiments: thus, although not prohibited, this network is not intended for carrying data traffic (i.e., traffic inside the experiments). Each of the iNodes is connected to an HP switch port over a Power-over-Ethernet (PoE) connection: this allows the central server to automatically remotely reboot the iNodes when starting a new experiment.

Within the OFELIA project, the purpose is to deploy on the iNodes OpenFlow switch implementation that switches the traffic sent over the wireless interfaces. These OpenFlow switches however will connect to (an) OpenFlow Controller(s) deployed on the iLab.t Virtual Wall. In this sense, the OpenFlow protocol messages will be sent over the wired network, that was intended for experiment life-cycle management and thus no guarantees on the quality of this connection can be provided.

It must also be noted that the HP switches on the different floors are not connected to each other and to the central w-iLab server through a physical LAN connection but through a VLAN (ID: 104) connection (sharing the same infra as other networks inside the building). Not shown on Figure 12, next to this central server also a file share server (wilabfs.test) is connected in a similar manner to this network, in order to store experiment specific data (i.e., logging info) by mounting the right directory on the iNodes. The w-iLab network makes use of address range 172.24.24.0/22.

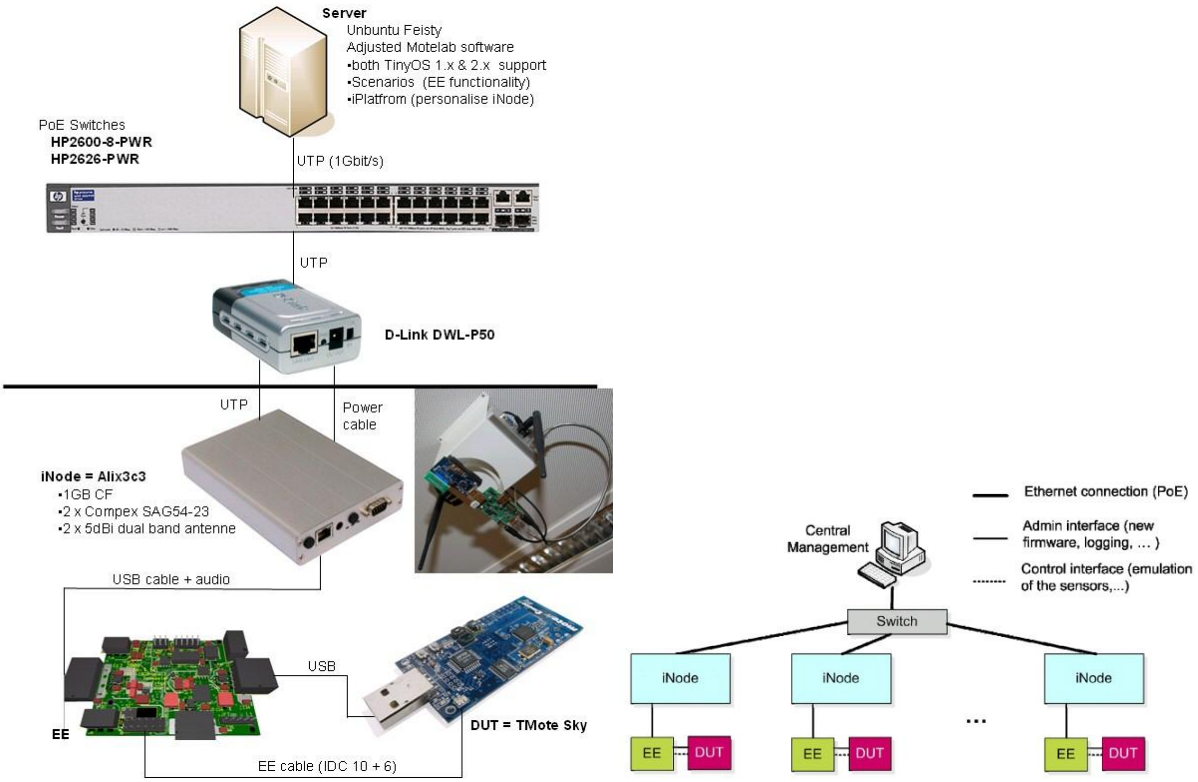

Figure 12: w-iLab.t network architecture

Since MS4.1, it was decided to migrate the w-iLab.t network to another address range (10.11.0.0/20) that is reachable from the outside world. However, the migration of each individual node to this new address range still needs to take place and is planned for the coming months. Currently only the w-iLab and wilabfs servers

have been migrated to their new address. This implies that while experiments are running, SSH login into the involved iNodes still needs to pass through the w-iLab central server on address 10.11.15.250 and with name wilab.atlantis.ugent.be as it acts as dynamic Network Address Translator (NAT) for which per experiment rules are configured based on personal IP addresses from which SSH connections can be established. More precisely, one should SSH into wilab.atlantis.ugent.be onto port 50000+<ID of iNode>. Also the wilabfs network share is still available (on its new name wilabfs.atlantis.ugent.be).

As mentioned above, the central w-iLab server is responsible for the experiment life-cycle management and providing a web interface through which researchers can define and control their experiments. Per experiment an iPlatform can be defined: such an iPlatform can specify which iNodes should mount what file shares, what scripts to start when booting and possible specifying a custom made kernel to be booted (loaded over PXE). It should be noted that, in contrast to the iLab.t Virtual Wall, no images are distributed to the iNodes at the start of each experiment: either the default Voyage Linux distribution available on the on-board CF card is booted or a custom distribution is mounted over the network. Experiments (jobs) get scheduled in chunks of 5 min blocks: scheduling an experiment implies reserving a complete zone for those 5 min blocks. To guarantee fairness, each experimenter gets only a certain quota in terms of minutes that he can schedule: once experiments are finished, the occupied quota is freed and becomes again for the user to be used for future scheduling. Scheduled experiments are stored in a database, against which a cron job checks every minute what experiments needs to be started and what needs to be stopped. As mentioned above, it is planned to replace this framework by an OMF based platform, which is also supposed to provide an SFA interface.

What is offered to the OFELIA project?

OFELIA and its users, will of course be allowed to register on the w-iLab.t and define and schedule experiments as every other w-iLab.t user, despite the fact that this mode of operation is not the one envisaged by the OFELIA project, that would have its own (expedient based) front-end.

Besides that, OFELIA is offered following two things:

- Initially, an SQL interface for filling in the database on the central w-iLab server to schedule jobs while bypassing the web interface was offered. However, this offering has slightly changed into a Python web scraping API that fills in the web forms as normal users are supposed to do. The web scraping approach has the advantage that no checks implemented in the w-iLab web server are overlooked and not implemented by another OFELIA w-iLab Aggregate Manager.
- Once the OMF-framework is deployed and if that indeed includes an SFA interface, OFELIA can use this interface.
- A physical port on the HP switches to connect the w-iLab network directly to the rest of the OFELIA infrastructure.

Constraints to be considered by the OFELIA project

Given the current deployment of the w-iLab.t test network, following constraints have to be taken into account:

- Assuming all OFELIA experiments will make use of the WiFi capabilities, experiments can only be deployed outside office hours (thus available from 8pm until 6 am and during the weekends), to protect IBBT personnel in the offices where the w-iLab.t testbed is deployed from excessive radiation.
- As w-iLab.t is a (time-)shared infrastructure, OFELIA and its users will be subject to a quota to be agreed upon.
- w-iLab.t still makes use of the internal 172.24.24.0/22 address range, preventing OFELIA of using this address range for other purposes, unless logical separation from the w-iLab.t infrastructure is realized and guaranteed. As mentioned above, migration to a public address range 10.11.0.0/20 is planned.
- VLANs are currently unavailable for isolating experiments from each other, as only too little entries in VLAN tables in the existing switch infrastructure are still unoccupied for other purposes.

- A DHCP service on w-iLab.t is configuring the LAN interface address on the iNodes.
- Logging in through the NAT functionality on the central w-iLab.t server requires specifying the address from which SSH connections are initiated.
- On the wired w-iLab.t network, no capacity or QoS can be guaranteed, as it is intended as a control network.
- On the iNodes, the default on-board Voyage Linux distribution should be used, or a customized one should be mounted over the network. In case OFELIA would develop a stable customized distribution (e.g., an OpenFlow capable OpenWRT distribution), this distribution may be programmed on the on-board CF cards in case sufficient storage capacity is left: such an operation can only happen once, requiring a sufficiently stable version.

3.2 IBBT island: operational report

Following issues/experiences were perceived, during the operations of the IBBT island.

- **Issue: OpenVPN connectivity interruptions**
  - **Component**: central hub infrastructure, in particular OpenVPN tunnels
  - **Description**: In phase I different islands are interconnected through OpenVPN tunnels. It was noticed that from time to time an island got disconnected while the others remained connected. After these outages, the automatic ping-restart feature of the OpenVPN service restored the connectivity.
  - **Solution / remediation**:
    - The root cause of the problem has not been identified; nevertheless, the situation seems to have improved somehow; therefore, the issue assumed to be closed, but further monitoring is going on.
    - The root cause of these outages is probably that the OpenVPN tunnels cross different networks relying on a best effort service. Thus temporary congestion on these networks or maintenance activities in these networks that we do not know about may have caused these kinds of interruptions in the service.
    - As initially the OpenVPN tunnels were only based on UDP, we have also reconfigured the OpenVPN server in the central hub to allow islands to connect over TCP rather than over UDP, as the unreliable service of UDP may aggravate problems in the TLS/SSL protocol used for the encryption. One island (the EICT island) has migrated to a TCP based OpenVPN tunnel, but it seems not the help solving the problem.
    - Most probably, the situation has improved as the ZenOSS monitoring system is continuously sending some traffic through the OpenVPN tunnels, keeping it naturally alive and reducing the impact (restarts) as more packets could potentially make it through a temporary congestion spot in one of the networks.
    - In Phase II dedicated lines will be leased and thus we should not rely anymore solely on the OpenVPN tunnels over best-effort network services.
- **Issue: link problems inside experiments on the Virtual Wall.**
  - **Component**: Virtual Wall
  - **Description**: When swapping out and swapping in a very basic experiment many times, intermittently problems occurred with some links in the experiment. In particular, two rare symptoms have been diagnosed: one symptom is that the 2 onboard interfaces available in the 60 nodes with 6 GbE interfaces may sometimes not get into an operational error-free state; the other symptom is that ARP broadcast messages may sometimes not get through the central Force10 switch.
  - **Solution / remediation**:
    - This is a critical issue that needs urgent remediation; investments are underway to resolve the problem.
    - Extra new network cards will be added to the machines with the 2 onboard interfaces and that have a free slot available (48 of the 60 machines). Further actions will be taken according to the evaluation of these new network cards.

        - For the other problem, several tracks are followed. Testing with OpenFlow experiments will be conducted on the second Virtual Wall: if successful, we may consider redirecting OFELIA experiments to that other Virtual Wall. As the central Force10 switch may not properly behave due to interference with LLDP packets often used in OpenFlow experiments (the way NOX performs topology discovery), protocols on the central Force10 switch will be shut down as much as possible to the bare minimum needed for the operation of the Virtual Wall. Finally, a firmware upgrade of the central Force19 switch is considered, requiring a new service agreement for this switch.
        - Users are recommended to swap out and swap in again the experiment in case they are confronted with this issue.
- **Issue: no connection to the public Internet**
    - **Component**: Virtual Wall and w-iLab.t
    - **Description**: The firewall around the IBBT test facility does not allow direct connection to the public Internet / external networks form the nodes on the respective infrastructures. This is an intentional IBBT policy and thus not a problem to be solved as such.
    - **Solution / remediation**:
        - Make use of port forwarding in the SSH tunnel, when logging in over SSH, or
        - If applicable (e.g., HTTP traffic), make use of the proxy server proxy.atlnatis.ugent.be
- **Issue: planned unavailability of infrastructure**
    - **Component**: w-iLab.t (and possibly Virtual Wall)
    - **Description**: On some occasions, it was requested on the w-iLab mailing list (not visible to others in OFELIA except the OFELIA IBBT Island Manager) not to use the w-iLab.t wireless testbed (and hence to free up any reservations in that timeslot), since it was needed for demonstration purposes. This was not a problem as such, since there were on those moments no OFELIA external users requesting access, but this issue is raised in anticipation of potential future conflicts.
    - **Solution / remediation**:
        - This issue is to be taken into account in case any reservation system will be developed in OFELIA.
        - Alternative, the aggregation manager interfacing with the w-iLab.t wireless testbed needs to get a feature to block reservations in these timeslots and to inform “owners” of reservations in that timeslot that they need reserve their experiment.

3.3 IBBT island: plans for Phase II

In accordance to the DoW, IBBT planned three things for Phase II: building a NetFPGA farm, enabling federated experiments and integration of the iLab.t Virtual Wall reservation system under development into the OFELIA control framework. Besides these three points, IBBT will also look into at one side migrating to SFA-based interfaces to the shared infrastructures and enhancing the currently provided OS images customized for the OFELIA project.

3.3.1 Building NetFPGA farm

IBBT will deploy and make available 10 NetFGPAs (see http://netfpga.org/) by the completion of the second phase. Such a NetFPGA board consists of a Xilinx Virtex-II Pro 50 FPGA chip and four times a GbE interface. These boards are plugged into a PCI slot. A NetFPGA allows programming a bitfile (so called gateware) in the FPGA chip. In OFELIA it is envisaged that the OFELIA users can program in hardware and experiment with their own OpenFlow switch implementations.

At the time of writing, the 10 NetFPGAs have already been installed into 5 servers. These cards will be wired to the central hub switch. Software installation is ongoing: the possibility for having a VM per NetFPGA is under consideration (e.g., virtualization based on XEN has already been shown as described in http://netfpga.org/foswiki/bin/view/NetFPGA/OneGig/NetFPGAVirtualizationWithXen).

Also an aggregate manager (AM) will need to be developed in order to interface with the OFELIA control plane. Initially, an Emulab based environment (cfr. http://www.protogeni.net/trac/Emulab/wiki/netfpga) was

envisaged. However, as the virtualization VT-AM developed in OFELIA makes use of XEN and as NetFPGAs can be made available to XEN VMs, an alternative solution building on the VT-AM is also under considered.

3.3.2 Enabling federated experiments

Enabling federated experiments across multiple islands, requires that the islands become interconnected. In first instance, an OpenVPN based L2 tunneling solution will be adopted. By phase 2, as part of the task T2.4 in WP2, dedicated lines between the different islands will be purchased. In case of IBBT as hub, a bundled service of 10 times a 1 GbE circuit will be leased over the BELNET network to Brussels: other islands are expected acquire a circuit to Brussels, where it will be patched onto one of the ten 1GbE circuits over BELNET. In Ghent, the last mile from the BELNET PoP to the IBBT premises will be implemented by means of a 10 GbE link: the exact multiplexing technique inside this 10 GbE circuit is still under discussion.

Besides this external connectivity, also the internal iLab.t Virtual Wall and w-iLab.t wireless sensor network needs to be connected to the central hub switch. For this purpose, on the Virtual Wall, a 10G link will be created between the OFELIA central hub switch and the Force10 switch in the Virtual Wall. The challenge here will be the coordination between the Emulab software controlling the Virtual Wall and the OFELIA Control Framework for assigning the proper VLANs on this link to separate experiments from each other.

The solution to connect the w-iLab.t wireless testbed to the OFELIA infrastructure is not yet decided. A port on one of the HP switches supporting the w-iLab control network could be made available for OFELIA. However, the w-iLab control network runs in a single VLAN, alongside some other networks in other VLANs. An additional constraint to take into account is the fact that only a limited number of VLANs can be configured on these switches, preventing us from using VLANs to separate individual experiments.

3.3.3 Integration of reservation system

As initially planned, IBBT still has the ambition to integrate its reservation system(s) into the OFELIA Control Framework, as we consider a reservation system to reserve the resources an experimenter needs to conduct his experiments an important feature of any test facility. However, due to several reasons (as a consequence that this feature is neglected in most control frameworks), we may decide in the end to drop this objective.

First of all, so far OFELIA does not feature a reservation system as it is providing a best-effort service. As long as this situation does not change, there is no point at all in integrating whatever reservation system that may exist on the Virtual Wall or w-iLab side.

Secondly, efforts to build a reservation system for the iLab.t Virtual Wall have so far not resulted in the expected outcome, as it turns out to be more complex than initially expected. At one side, the system needs to deal with heterogeneous resources (some nodes have 4 other 6 pots; some nodes have a screen attached while others don't) which involve some optimization tasks, while at the other side the system needs to be integrated in the Emulab software.

Thirdly, as a migration towards an OMF based control framework on the w-iLab.t wireless sensor network is going on, its current reservation system will become obsolete and so far it is unclear how this issue will be tackled on the w-iLab infrastructure itself.

3.3.4 Migration to SFA-based control interfaces

Currently the expedient plug-in interfacing with the Emulab software on the iLab.t Virtual Wall makes use of the Emulab proprietary xml-rpc interface. As latest versions of the Emulab software features a ProtoGENI interface, which is SFA based, our intention is to migrate to this ProtoGENI interface. However, this turned out to be unsuccessful during Phase I; we will continue this effort for Phase II.

Also as the control framework of the w-iLab.t wireless testbed is being migrated to an OMF based control framework, it is supposed to expose an SFA based interface since OMF appears to feature such interface. This interface will be investigated in more details and when possible the necessary changes in the proper expedient plug-in will be made in order to make use of this SFA based interface.

3.3.5 Enhancing OS images customized for OFELIA

In phase I, for the iLab.t Virtual Wall and the w-iLab.t wireless testbed an OS image has been prepared and customized specifically for the OFELIA project. A new image will be created for by the end of Phase II, featuring more software packages (e.g., besides OpenFlow 1.0 also OpenFlow 1.1 software; all kinds of tools like Wireshark that network researchers might find useful; besides the NOX controller other OpenFlow controllers; etc) and an upgrade to more recent versions of the software packages already supplied in the current OFELIA OS images.

4 ETHZ island

4.1 ETHZ island: current Status

4.1.1 Topology and connectivity

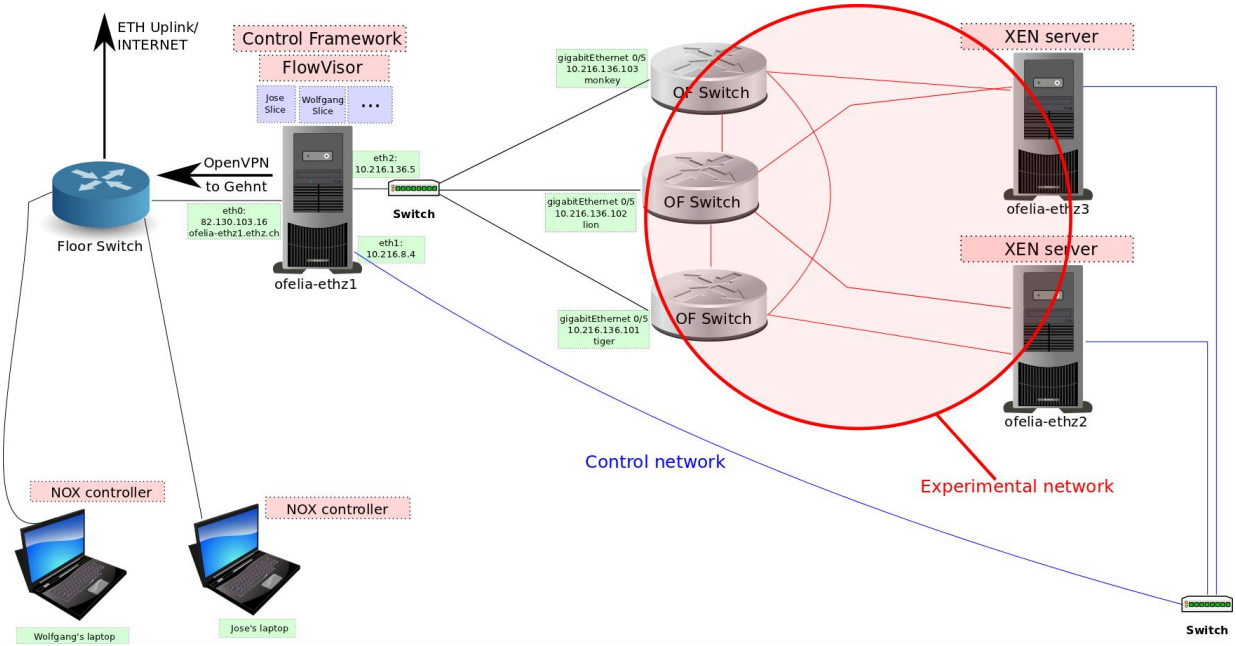

Figure 13: ETH Zurich island topology

The island topology is shown in Figure 13. In our island the logical topology of three different networks (experimental, control, management) can be found mapped in also three different physical networks.

The experimental network topology is highlighted in red. Notice that the physical links shown in the experimental network are not meant to be definitive. Our goal is to converge to a topology which shows to be the most useful one according to the requirements of experimenters. In that sense the challenge is to link our servers to the switches so that a rich number of virtual topologies are possible. Currently we run a full-mesh topology between the switches.

The control network gives access to administrate the XEN servers, while the goal of the management network is to provide connectivity between FlowVisor and the OpenFlow switches. Also notice that in addition to administrate the host XEN machines, the control network can serve as a path towards FlowVisor for controllers who are sitting in the XEN VMs.

In order to save interfaces in our ofelia-server1 we connect it to the control network and management network using dumb L2 1GBit switches. Notice that in control network and management network we rely on IPv4/MAC for communication and no experimentation traffic is expected.

So far we have tested to run our controllers externally and also within the XEN servers, both solutions seeming to be compatible.

4.1.2 Hardware Description

Inventory of OpenFlow switches			
Manufacturer	**Model**	**Management IP**	**Mode of operation**
NEC	NEC IP8800/S3640-24T2XW	10.216.136.101	RSI
NEC	NEC IP8800/S3640-24T2XW	10.216.136.102	RSI
NEC	NEC IP8800/S3640-24T2XW	10.216.136.103	RSI

Inventory of servers				
Architecture	**OS (host)**	**RAM**	**Networking interfaces**	**Duties**
64-bit	Debian Squeeze 64-bit	36 GB	4 Ethernet x 1GBit	Control Framework, FlowVisor
64-bit	Debian Squeeze 64-bit	36 GB	4 Ethernet x 1GBit	XEN VMs
64-bit	Debian Squeeze 64-bit	36 GB	4 Ethernet x 1GBit	XEN VMs

Additional less relevant hardware: 2 L2 1Gbit switches, 1G Ethernet cables.

4.1.3 Access to the island

Our island has connectivity through the OpenVPN tunnel to Gent. Our OpenVPN software is running in our ofelia-server1 machine. In addition, this machine has the necessary routing table to provide connectivity to the control network from to the other islands. Reachability from/to other islands control network has been tested successfully.

4.1.4 Experimenters in the island

Currently two students are running some experiments and participating as alpha-testers of our deployment. Experimentation in the island seems to be working; however it is not possible to completely set-up full experiments using the control framework due to its instability still present.

Also to notice, the experiments that our students wanted to run very often required close access to the switches. It means, low delays, extraction of relevant information of the rules in the switch, etc. However, it turned out that FlowVisor interfered in the experiments much more than thought in the beginning. An example of this is the rate limitation of FlowVisor for packet-in messages. While for a regular operation this can be a nice feature, when using experimentation it might interfere undesirably.

4.2 ETHZ island: operational report

Following operational issues were encountered in the ETHZ island:

- In ETHZ island we fixed the refreshment of the topology mechanism. We tested it successfully for a while. However, FlowVisor developers changed the API of FlowVisor, making impossible the refreshment topology mechanism we had just fixed. In the meantime we contacted with Rob Sherwood (one of the FlowVisor developers) and the API was partially reverted and we could use again the topology refreshment mechanism. However, now the FlowVisor API has changed again, and the topology refreshment becomes again unstable. Therefore, it seems that we must think how to not depend on FlowVisor API since there are significant risks that it will change unexpectedly. In this context, the changes could come not only for the topology refreshment but for any other usage (i.e. reserving flowspaces or removing them from FlowVisor).
- Several students performed a number of experiments in the testbed. During these experiments the students experienced diverse technical problems. Most of those problems can be related to the same source. The students have problems in the communication controller-switch because of FlowVisor. Although in theory FlowVisor should behave transparently for both controller and switches, in practice it is not the case. Therefore, debugging an experiment, controlling how many rules are installed in the switches or any other relatively simple tasks can become quite complicated or even impossible. In that direction we found that FlowVisor made a gap between the programmable OpenFlow hardware and the control plane.
- We found also quite tedious the set-up of the virtualization agents within the virtualization manager. Although it must be done only once, it is quite error-prone and in case some XEN servers are changed this configuration must be changed again. We believe much of the information needed to set-up a virtualization agent within the virtualization manager could be guessed automatically by the

agent, or other components or the framework, and we wonder why some of the parameters are required.

4.3 ETHZ island: plans for Phase II

4.3.1 Approaching the integration of real users in the testbed

One of the goals of ETH Zürich island is the possibility of use real users' traffic in the testbed. By real users we mean people who are using the network for its daily production work. An example of this could be students in the faculty, or other researchers of the group that are browsing the web, checking their email, etc.

We conceive to main challenges/steps to enable the integration of real users' traffic in the testbed:

- Give the real users connectivity to the testbed: the way real users will inject their traffic in the testbed can be very different from one approach to another. We envision several possibilities:
    - The simplest one is to just plug the cable of the real user directly in one of our OpenFlow switches in the testbed. However this has some drawbacks. First the real user does not have a backup connection. Second, the real user only contributes his traffic when he is sitting in the location of the cable which gives him access to the Internet.
    - A second approach, however more tricky and probably higher to manage is to deviate the traffic of users connected to our flow switch through the testbed. Here in our institute all users are connected to our floor switch to get daily access to the Internet. Users are identified by their MAC addresses. If we configured the floor switch to deviate all the traffic coming from/going to a specific MAC address to pass through the testbed, then this traffic could be used for experimentation. Of course the drawbacks of this approach include a pre-configuration of the floor switch and the need for re-configuring it in case something goes wrong and the real users requires accessing the Internet. There is also a risk of misconfiguration of the floor switch, which is a high risk since our entire institute members, including professor, researchers, students, etc. are connected to the switch.
    - A more flexible approach would be as follows. We provide real users credentials to connect to the VPN endpoint of our island. The real users connect to the VPN wherever they are (at home, at work, on holydays) and whenever they wish their traffic is carried through the experimental network. The advantages of this approach are clear. The real user can contribute his traffic from wherever he is. In case he has trouble with the experiment, he just needs to disconnect from the VPN and will be able to use its regular connection. One additional overhead is that the CF should take care of who connects to the VPN and in which experiment opts-in.
- Give the testbed connectivity to the Internet: the experimental network should have access to the Internet. In the end, if we want the traffic of real users to flow across our island, we need to give the real user the possibility to reach the Internet. If we did not offer access to the Internet from the testbed the real users will lose any motivation they could have to participate in experimentation. This is because most of the services real users want to use rely to a big extent on the connectivity to the rest of the Internet and not only our testbed. Furthermore, we believe that according to our current topology, the most convenient part of the network to "connect" to the Internet is the experimental network.

  A good approach to accomplish this task would be as follows. We connect a traditional NAT box in our experimental network. This NAT box has Internet connectivity, as well as a public IP. Then, the private IP of the NAT box that the experimenters must use as a gateway can be reserved also through the CF. Ideally the UI of the CF would include a checkbox with which the experimenter can ask for Internet connectivity. Then, this would be translated to an extra portion of the flowspace reserved to that slice, so that the experimenter can send and receive traffic from the NAT box. We are aware that there are many details and limitation with this approach. However, we believe it is a good solution for a first step. Indeed, we consider this solution very secure, as we use the NAT box also as a "firewall" to disallow possible unwanted incoming traffic from the Internet. We assume that all the traffic willing to go outside the facility towards the Internet must be IP traffic. Specifically, due to the restriction of the commercially available NAT boxes, the traffic must be TCP, UDP, or IMCP.

We are still studying which of these possibilities would be the more convenient according to its advantages, disadvantages and deployment time required.

4.3.2 Provide traffic generators

In practice it might turn out that many experimenters reserve virtual machine with the only goal of generating some traffic. We believe this is a waste of resources of the facility and makes more tedious the set-up of the experiments for the experimenter.

We envision the possibility to have a specific sort of resources that are in charge of generating traffic according to some settings of the experimenter. In this settings could be, for instance: length of the packets, frequency to send the packets, structure of the frames (MAC?), and headers content (should be check that the requested headers fit in the experimenter slice).

In the implementation side, these traffic generators could be just a specific type of VM. Therefore, there would be a template specific that the experimenter can choose when requesting a VM. Instead of creating a XEN VM the control framework will put the traffic generator to run. How the traffic generator should look like is something still under study.

4.3.3 Study possible federation with GpENI

We are still looking at how a federation with GpENI would be possible. Not only on the technical side but also political concerns as well as security must be kept in mind.

4.3.4 Encapsulate network components in “management network”

One of the goals of the “management network” is to keep hidden in each island network elements that should not be exposed publicly. An example is the communication channel between FlowVisor and the OpenFlow switches. As so, our island currently encapsulates as much as possible.

Nevertheless we are trying to figure out how to solve the drawbacks coming from this. We noticed that, in order to allow ZenOSS to reach the OpenFlow switches, ZenOSS needs to reach them. This contradicts the idea of isolating the management network from the control network.

After trying different possibilities we will come up with the most convenient solution.

5 I2CAT island

5.1 I2CAT island: current Status

In this section we will describe the high speed network that i2CAT has available and also the current infrastructure that i2CAT has deployed for OFELIA facility testbed. It currently consists in only L2 (Ethernet) switches and computational substrate (servers).

5.1.1 Equipment specifications.

5.1.1.1 Network equipment

The island has five NEC switches model IP8800/S3640-24T2XW. This equipment neither will use SFP nor XFP optical transceivers. Only copper ports will be used. The following table details the specification of the purchased model (24-port model).

		S3640-24T2XW
Maximum Switching Capacity		88 Gbps
Maximum Packet Processing Performance		65.5 Mpps
Network Interface	1000BASE-T	24 (*1)
	1000BASE-X (SFP)	4 (*1)
	10GBASE-R (XFP)	2
MAC Address Table		32768 entries
VLANs		4094
Form Factor		1U

*1: Four ports can be configured as SPF ports

Figure 14: Table with NEC switch specifications

In addition, 3 HP E3500-48G-PoE+ yl OF-Enabled switches are purchased and will be installed in phase II. The OF-enabled switches are running version K.14.79o OF 1.0 compliant.

5.1.1.2 Servers

There are 5 dedicated servers:

- 3 Supermicro SYS-6016T-T NEHALEM E5506 2,13 Ghz 12Gb RAM 2TB disk (2u)
- 2 Supermicro SYS-6016T-T WESTMERE 2,4 GHz 12Gb RAM 2TB disk (2u)

5.1.2 Inventory

Inventory of OpenFlow switches in the testbed			
Manufacturer	**Model**	**Datapath ID**	**Status**
NEC	NEC IP8800/S3640-24T2XW	00:10:00:00:00:00:00:01	Up and running
NEC	NEC IP8800/S3640-24T2XW	00:10:00:00:00:00:00:02	Up and running
NEC	NEC IP8800/S3640-24T2XW	00:10:00:00:00:00:00:03	Up and running
NEC	NEC IP8800/S3640-24T2XW	00:10:00:00:00:00:00:04	Up and running

NEC	NEC IP8800/S3640-24T2XW	00:10:00:00:00:00:00:05	Up and running
HP	HP E3500-48G-PoE+ yl	-	Unavailable
HP	HP E3500-48G-PoE+ yl	-	Unavailable
HP	HP E3500-48G-PoE+ yl	-	Unavailable

Inventory of servers in the island					
Model	**Host name**	**Operating System**	**RAM**	**Duties**	**Status**
SuperMicro SYS-6010T-T	Llull	Debian Squeeze 64-bit	12 GB	Control Framework, FlowVisor, VPN tunnel	Up and running
SuperMicro SYS-6010T-T	Foix	Debian Squeeze 64-bit	12 GB	XEN Server	Up and running
SuperMicro SYS-6010T-T	March	Debian Squeeze 64-bit	12 GB	XEN Server	Up and running
SuperMicro SYS-6010T-T	Rodoreda	Debian Squeeze 64-bit	12 GB	XEN Server	Up and running
SuperMicro SYS-6010T-T	Verdaguer	Debian Squeeze 64-bit	12 GB	XEN Server	Unavailable

5.1.3 Topology and network configuration

The current topology is composed by the five NEC switches and the five servers. The remaining three HP switches are racked and have connection to the control and management network, but are not yet operational.

The topology deployed can be shown in Figure 15:

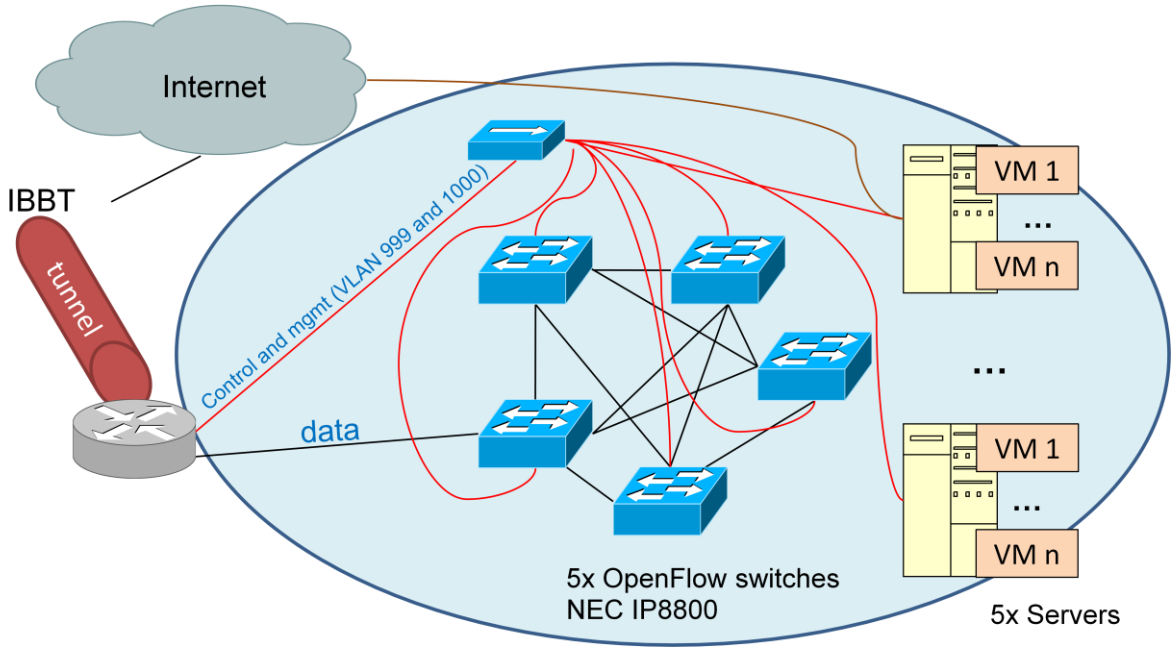

Figure 15: Current operational i2CAT island topology for OFELIA facility (HP switches not yet in production)

The network is split into three logical networks, in a similar way than other Ofelia islands, with out-of-band management. NEC switches are used in RSI (Real Switch Instance), and are configured with ports 1-16 to use OpenFlow and the remaining 8 ports to use legacy protocols (trunk mode for VLANs 999 and 1000):

- **Control network (VLAN 999):** this network is used for interconnect VPN and XEN VMs, to allow users to connect through SSH to the VMs or access local Control framework instance (Web portal).
- **Management network (VLAN 1000):** this part of the network is used for management. It interconnects servers and control framework virtual machines (located in Llull server) as well as the management interfaces of the switches.
- **Data network (OpenFlow network):** Is used for experimentation. Each of the servers has two connections to two of the NEC switches. These interfaces are shared between VMs through a bridge (in this case, VMs have eth1 and eth2 connected to the OpenFlow network). See table below for details.

Switch	Datapath ID	Port	Server	Virtual Machine Interface
NEC1	00:10:00:00:00:00:00:01	11	Llull	eth2
NEC1	00:10:00:00:00:00:00:01	12	Verdaguer	eth2
NEC2	00:10:00:00:00:00:00:02	11	Llull	eth3
NEC2	00:10:00:00:00:00:00:02	12	Verdaguer	eth3
NEC3	00:10:00:00:00:00:00:03	11	Foix	eth2
NEC3	00:10:00:00:00:00:00:03	12	Rodoreda	eth2
NEC4	00:10:00:00:00:00:00:04	11	Foix	eth3
NEC4	00:10:00:00:00:00:00:04	12	March	eth2
NEC5	00:10:00:00:00:00:00:05	11	March	eth3
NEC5	00:10:00:00:00:00:00:05	12	Rodoreda	eth3

Figure 16: OpenFlow domain connections switches - servers

Switch	Port	Connected to	Switch	Port
NEC1	2		NEC2	1
NEC1	3		NEC3	1
NEC1	4		NEC4	1
NEC1	5		NEC5	1
NEC2	3		NEC3	2
NEC2	4		NEC4	2
NEC2	5		NEC5	2
NEC3	4		NEC4	3
NEC3	5		NEC5	3
NEC4	5		NEC5	4

Figure 17: OpenFlow domain connections between switches

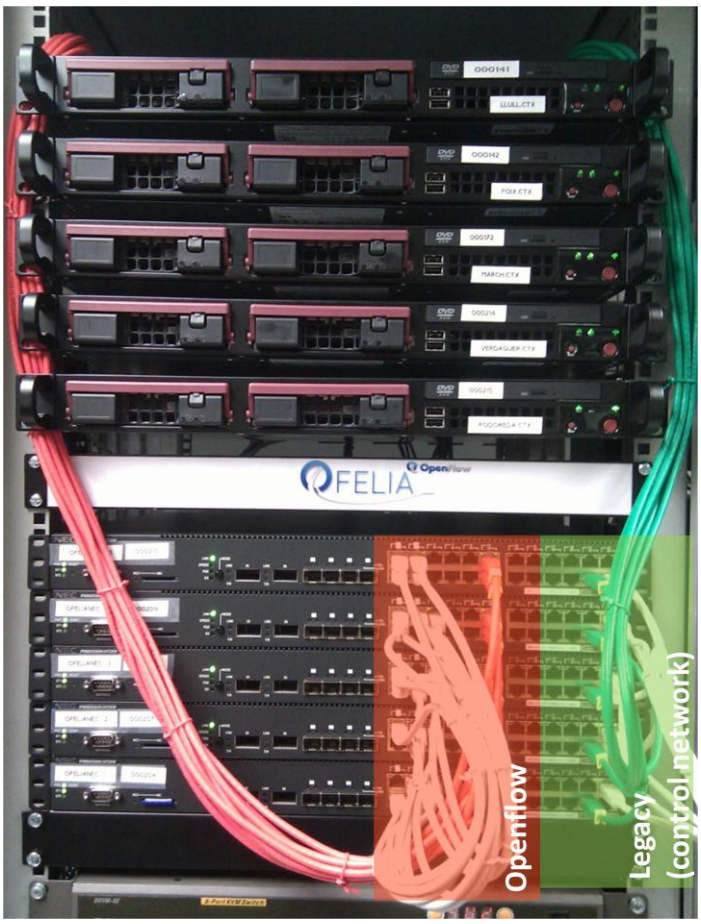

Figure 18: Equipment in place (NEC switches only)

5.1.4 Addressing

Following OFELIA's addressing convention the following ranges are being used at the moment:

- **Control network (VLAN 999):** 10.216.12.0/22
- **Management network (VLAN 1000):** 10.216.140.0/22
- **Data network (OpenFlow network):** No restriction

The currently allocated addresses are discussed in the following two subsections.

5.1.4.1 Control plane traffic (VLAN=1000)

IP range: (10.216.140.0 - 10.216.143.255)

IP / IP range	Machine	Type	Alive
10.216.140.1	Llull	Assigned	Yes
10.216.140.2	Foix	Assigned	Yes
10.216.140.3	March	Assigned	Yes
10.216.140.4	Verdaguer	Assigned	No
10.216.140.5	Rodoreda	Assigned	Yes
10.216.140.6		FREE	
10.216.140.7	CF VM	Assigned	Yes
10.216.140.8	Database VM	Assigned	Yes
10.216.140.9	FlowVisor 1	Assigned	Yes
10.216.140.10	FlowVisor 2	Assigned	No

10.216.140.11	PreProduction CF	Assigned	Yes
10.216.140.11-20.216.140.20		RESERVED (New Hosts)	
10.216.140.21	NEC1	Assigned	Yes
10.216.140.22	NEC2	Assigned	Yes
10.216.140.23	NEC3	Assigned	Yes
10.216.140.24	NEC4	Assigned	Yes
10.216.140.25	NEC5	Assigned	Yes
10.216.140.26	HP6	RESERVED	No
10.216.140.27	HP7	RESERVED	No
10.216.140.28	HP8	RESERVED	No
10.216.140.29-10.216.140.50		RESERVED (New Switches)	
10.216.12.51 - 10.216.143.255		FREE	

5.1.4.2 User plane traffic (VLAN=999)

IP range: 10.216.12.0/22 (10.216.12.0-10.216.15.255)

IP / IP range	Machine	Type	Alive
10.216.12.1	CF VM	Assigned	Yes
10.216.12.2	Local LDAP	Assigned	No
10.216.12.3, 10.216.12.4	FlowVisor 1 and 2	Assigned	Yes, No
10.216.12.5	Llull (gateway Ofelia and Internet)	Assigned	Yes
10.216.12.6 - 10.216.12.25		RESERVED (Local services)	-
10.216.12.26 - 10.216.15.255	Users VMs	FREE	

5.2 I2CAT island: operational report

Following operational issues were encountered in the I2CAT island:

- **Issue: OpenVPN connectivity interruptions**
  - **Component**: central hub infrastructure, in particular OpenVPN tunnels
  - **Description**: In phase I different islands are interconnected through OpenVPN tunnels. It was noticed that from time to time an island got disconnected while the others remained connected. After these outages, the automatic ping-restart feature of the OpenVPN service restored the connectivity.
- **Issue: Extremely slow performance on framework components.**
  - **Component:** All
  - **Description:** During first days of operation of i2cat's island it was observed that, without any change in neither the configuration nor the software, suddenly the web applications were performing very slow (lasting up to 3 minutes to load a page).
    Investigation revealed that the problem was that the MySQL engine was trying to make a reverse lookup for each and every external connection (i2cat's island uses 1 separate machine for MySQL server). The problem was that the only DNS (wrong configuration) set in */etc/resolv.conf* file was down, hence not being able to perform reverse lookups.
  - **Solution**: To solve this issue:
    Add appropriate DNSes under */etc/resolv.conf*; primary and secondary.

    In addition, MySQL engine was configured to avoid reverse lookups (it can increase a little bit performance on external MySQL queries). Adding the following at the end of */etc/mysql/my.cfg*:

    ```
    #Skip name resolv
    skip-name-resolve
    ```

Remember that this change will only take effect after MySQL engine restarts

- **Issue: Permission requests are not received.**
  - **Component:** Expedient
  - **Description:** Even the ROOT_EMAIL in the localsettings.py file was properly set, when a user was asking for permission to create a new Project, no notification arrived to the mailbox.
  - **Solution:** This behavior was caused by the fact that this feature was not developed. In v0.12 of the OFELIA Control Framework (OCF) this misbehavior was corrected. The configuration required is located in the *localsettigs.py* in Expedient:

```
EMAIL_HOST = "smtp.gmail.com"
EMAIL_USE_TLS = True
EMAIL_HOST_USER = 'OpenFlow@gmail.com'
EMAIL_HOST_PASSWORD = 'password'
EMAIL_PORT = 587
DEFAULT_FROM_EMAIL = 'no-reply@fp7-ofelia.eu'
EMAIL_SUBJECT_PREFIX = '[OFELIA CF]'
```

    In the request form, the list of available users to ask for permission is composed by the IMs (is_superuser = 1). So, the email to which the request will be sent is the superuser account, set in the installation process when synchronizing the database for the first time, and configurable in the Account (*https://expedient-url/users/detail/)* section when logged with this user.

- **Issue: 'Directory not empty' error**
  - **Component :** OXA (Agent)
  - **Description:** If users experience this error during VM creation:

```
Action create on VM test failed: : [Errno 39] Directory not empty:
'/tmp/oxa/hdtest_3382/'
```

  - **Solution:** It is most likely due to a wrong configuration of the server, especially /etc/modules file. Please, revise XEN installation manual and note that loop module **must** have max_loop=64 or higher.

```
root@node04:/etc# cat modules
# /etc/modules: kernel modules to load at boot time.
#
# This file contains the names of kernel modules that should be loaded
# at boot time, one per line. Lines beginning with "#" are ignored.
# Parameters can be specified after the module name.

8021q
loop max_loop=64
```

    Remember that changes in /etc/modules will not take effect until the system is rebooted.

- **Issue : VM access and actions in the OFC involving LDAP failed**
  - **Component:** All
  - **Description:** When trying to perform actions where information stored in the LDAP (projects, slices, and users) was required, the CF failed. In addition, accessing to a VM created in the project was denied although a username and password of a user belonging to the project were used.
  - **Solution:** The problem was caused by the DNS. It is important the first DNS in */etc/resolv.conf* to be the OFELIA internal DNS (currently 10.216.24.2), since it us used to resolve the IBBT's LDAP IP. This internal DNS also resolves external queries via gateway. Add 10.216.24.2 as the primary DNS in */etc/resolv.conf*

5.3 I2CAT island: plans for Phase II

5.3.1 Extension of the network topology

So far the Ethernet substrate of the island was composed by 5 NEC IP8800/S3640-24T2XW switches. In order to extend it 3 HP E3500-48G-PoE+ yl OF-Enabled switches has been acquired. It has still to be decided, but most probably the introduction of these new switches will maintain the completely mesh topology of the island. Specific connections between switches and servers have to be decided.

The 3 HP E3500-48G-PoE+ yl OF-Enabled switches will be using only copper ports for the moment. The main specifications can be seen in table below:

Table 4: HP switches specifications

Technical specifications	
Ports	1 open module slot; 44 autosensing 10/100/1000 ports(IEEE 802.3 Type 10Base-T, IEEE 802.3u Type 100Base-TX, IEEE 802.3ab Type 1000Base-T), Media Type: Auto-MDIX, Duplex: 10Base-T/100Base-TX: half or full; 1000Base-T: full only; 1 RJ-45 serial console port; 4 dual-personality ports, each port can be used as either an RJ-45 10/100/1000 port (IEEE 802.3 Type 10Base-T; IEEE 802.3u Type 100Base-TX; IEEE 802.3ab 1000Base-T Gigabit Ethernet) with PoE or an open mini-GBIC slot (for use with mini-GBIC transceivers); Supports a maximum of 4 10-GbE ports
Memory and processor	10G Module : ARM9 @ 200 MHz, packet buffer size: 36 Mb QDR SDRAM; Management Module : Stackable memory and processor: Freescale PowerPC 8540 @ 666 MHz, 4 MB flash Mb, 128 MB compact flash, 256 MB DDR SDRAM
Latency	1000 Mb Latency: < 3.4 µs (FIFO 64-byte packets); 10 Gbps Latency: < 2.1 µs (FIFO 64-byte packets)
Throughput	up to 111.5 million pps
Routing/switching capacity	149.8 Gbps
Switch fabric speed	153.6 Gbps
Routing table size	10000 entries
Management features	HP PCM+; HP PCM (included); command-line interface; Web browser; configuration menu; out-of-band management (serial RS-232C)

Addressing of the switches will be as depicted in the table below:

IP	Machine	Type	Alive
10.216.140.26	HP6	RESERVED	No
10.216.140.27	HP7	RESERVED	No
10.216.140.28	HP8	RESERVED	No

5.3.2 Perform connectivity tests between i2CAT and UEssex.

As first step in order to finally federate all the different islands, i2CAT will start testing the deployment of a link to UEssex's island. Probably both cases will be tested: through a dedicated VPN or through GEANT.

In addition these testing will be the former step to deploy the dedicated federation gateway of the island and to apply required changes to the different equipment already installed.

5.3.3 Increase island performance (Optional)

It will be studied the case of providing the island with major performance. To accomplish this goal one possible way would be avoid using central IBBT LDAP and VPN servers, but deploying them locally. In this way it would be assured the connectivity and operation of the island in case the central island in IBBT suffered any problem. Furthermore, by installing local servers information becomes redundant across the islands.

5.3.4 Deployment of local monitoring tools (Optional)

Linked to the optional plan above, it will be considered to deploy some monitoring system acting locally in the island. ZenOSS Service Dynamics appears as a possibility since it is already used in OFELIA.

5.3.5 Replace the Linux software bridge by OpenVSwitch instances (Optional).

So far network configuration inside the virtualization servers is based in Linux soft bridge software. As it has been repeatedly discussed in several project meetings, the option of replacing these bridges with OpenVSwitch will be studied. Since OpenVSwitch offers much more capabilities than Linux bridges it would provide better control between different virtual interfaces of users VMs and the physical cards in the server providing better isolation and slicing of the traffic.

6 TUB island

6.1 TUB island: current Status

In the OFELIA-TUB-Island is the target to provide an OF networking infrastructure for the local networking researchers and students by replacing the current switches with OF compatible HW. Therefore the next sections should show the current installation status which is an improvement of the basic installation looks like.

6.1.1 Topology and connectivity

The topology is separated into the physical structure and the logical connectivity. After functionality validation of the facility topologies real user traffic (researcher and student) should be handled by the OF infrastructure. At the moment only two end-hosts are connected for the validation process.

6.1.1.1 Physical connectivity

The physical connectivity issue is the wiring of the TUB's OF developers and student rooms. The switches which are connected with these rooms are located in the TUB-FR campus building at wiring-centre FR5539 and FR5048 in the fifth floor. The following picture shows the current physical topology of the wired switches. The red lines are 1GBit fiber cables and the green are 1GBit copper cables. We used a meshed network structure to get a redundant wiring and the possibility for different OF experiments.

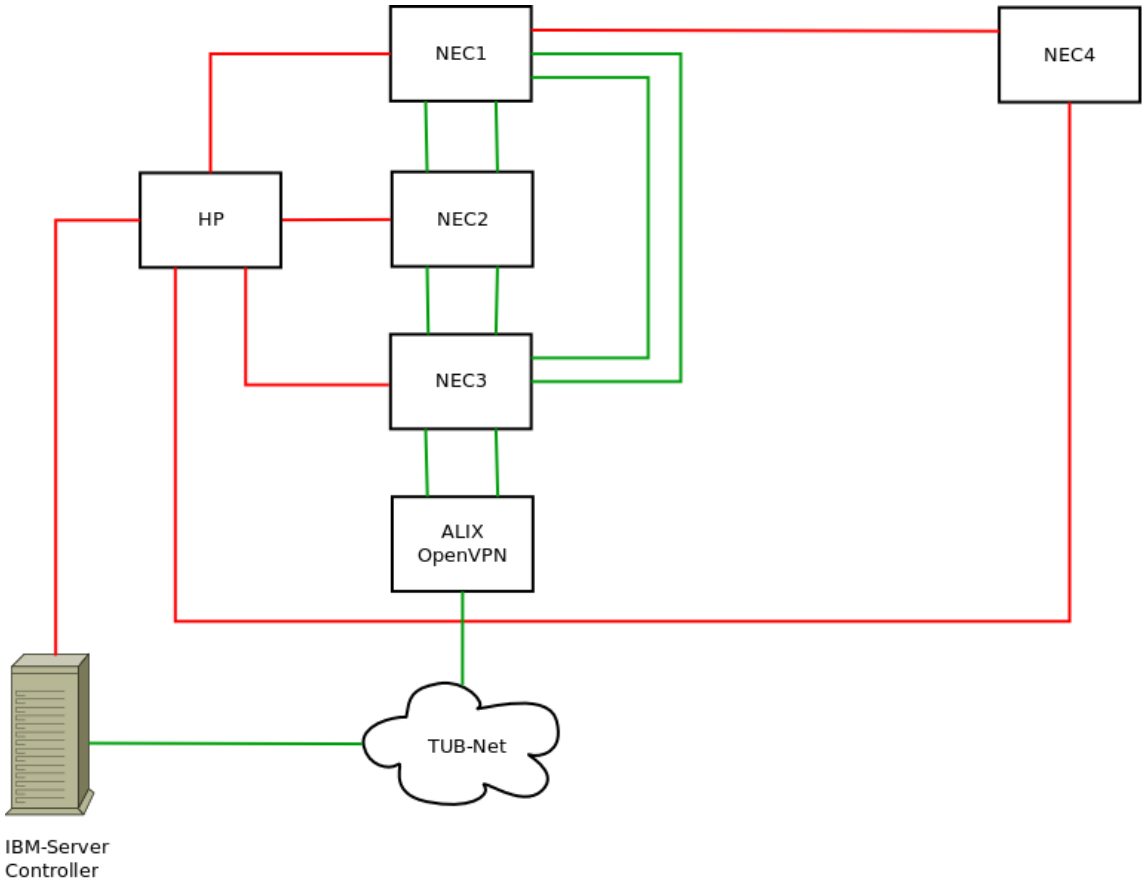

Figure 19: Physical topology

Figure 19 shows the physical topology of the improved installation in FR building with the OF compatible HW.

6.1.1.2 Logical topology

The physical available Ethernet links are logically separated into standard configured and OF enabled connections. It is planned to separate the links/ports workgroup dependent with OF slicing techniques, e.g. in a simple case by using different VLAN's including an associated controller. The OF developer- and research groups will be connected to OF-Ethernet data links. Therefore the connected PC's get IP addresses from the official TUB pool. Packets from these PC's will be switched by OF-equipment to the local router and then routed through TUB's standard networking infrastructure. For the switch administration and connection to the OF-Controller-Proxy (FlowVisor) is used a dedicated legacy control VLAN and a User-Control VLAN, too. Remote access and the option to use a custom controller will be given through a SSH connection with the IBM-Server which is mounted in the EN-Data-Center.

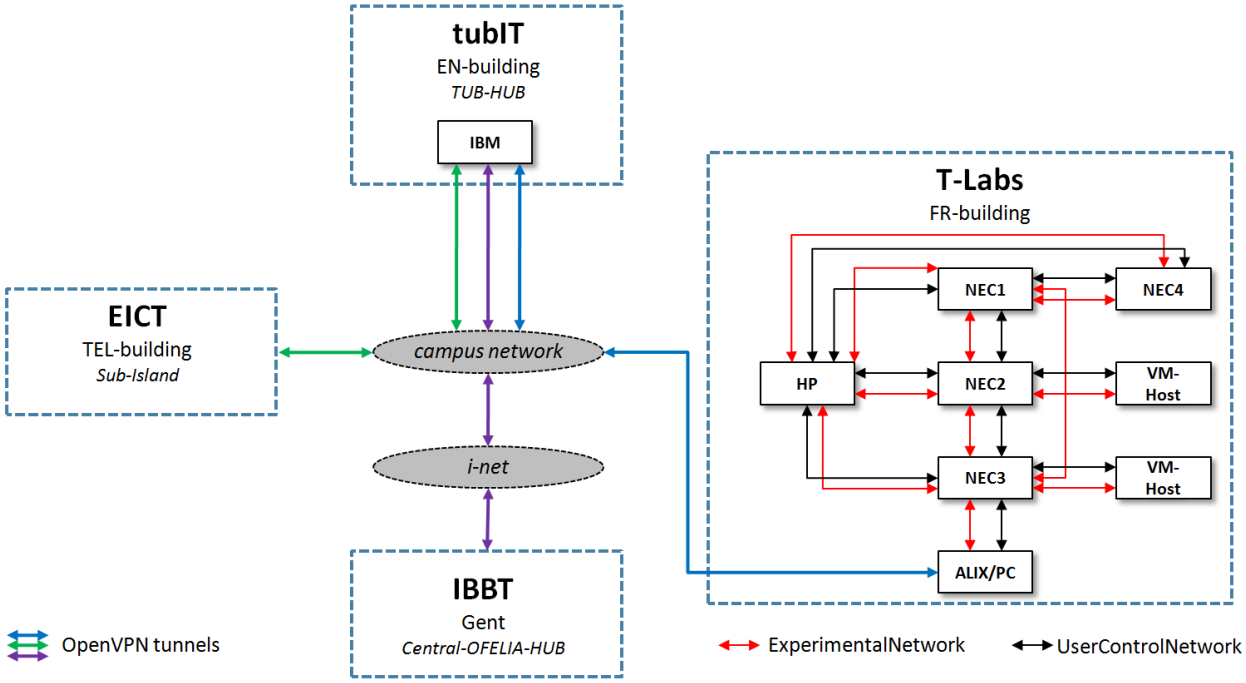

Figure 20: Sub-Island connection

Figure 20 gives an overview about the current connection status and geographical distribution of Berlin's sub-islands.

Island parameters in facts:

- Management Network
  - dedicated connection between switch and IBM-Server
  - NEC switch internal VLAN id 2
  - IP subnet range 10.0.0.0 / 24
  - Switch management (only TUB-Island internally used)
- User Control Network
  - OpenVPN tunneled
  - NEC switch internal VLAN id 3
  - IP subnet range 10.216.16.0 /22
- Experimental Network
  - OpenVPN tunneled
  - Dedicated NEC switch wiring

6.1.2 Hardware configuration and setup

The configuration and setup of the hardware will be described in this sub-section. It is separated into two parts, the switches which are installed in the FR-building and IBM-Server which is installed in the EN-building into the TUB's data-centre. These are the main setup components for supporting OF-Networking in the context of the OFELIA-TUB-Island concept.

6.1.2.1 Installed Switches

Number:	Product:
4	NEC IP8800/S3640-48TWLW with 48 10/100/1000 BASE-T LAN interfaces and 4x SFP
1	5400 with a 24 port SFP module with 16x HP SFP MM-SX duplex transceivers
1	ALIX embedded OpenVPN Tunneling-Device

Table 5: Installed switches

The four NEC switches are used as meshed access-switches and were connected over the patch-panel to the user-pc with Cat. 5U/UTP cables, which are supporting the 1000BASE-T networking standard.

The HP5400 switch is used as a segregation-switch and is connected with all the four access-switches. The switches are wired with duplex short wave multimode fibers to the HP-switch. This switch has a HP business swap guarantee in case of defect.

The ALIX Device is not a switch. It is an embedded pc with a SSH- and Web-Management-Interface required for the Layer two OpenVPN bridging of the User-Control-Network and the Experimental-Network. This device is directly connected with the untagged two network ports mentioned before and to the TUB campus-net where the tunnel connection is working with.

6.1.2.2 IBM-Server

The rack mountable IBM Server is a multipurpose device and should have enough performance and potential to add computing power to handle the different und maybe rising usage requirements. For now we use this device as the local tunneling hub and as Xen-Host. For example FlowVisor is hosted there.

Number:	Product:
2	Intel Xeon 4C Processor Model E5620 80W, 2.40GHz/1066MHz/12MB
3	IBM Memory 4GB (1x4GB, 2Rx4, 1.5V) PC3-10600, CL9, ECC, DDR3, 1333MHz
2	IBM 146 GB 2.5" SFF Slim-HS 10K 6Gbps SAS HDD
1	PRO/1000 PF Server Adapter by Intel with SR LWL Port, GbE
1	IBM Emulex 10GbE Virtual Fabric dual-port Adapter
2	IBM/Brocade 10Gb SFP+SR Optical Transceivers
1	IBM Virtual Media Key extension of the System for Remote Console View
2	IBM 675W Power supply
1	IBM UltraSlim Enhanced SATA Multi-Brenner
1	Linux Enterprise Server 1-32 Sockets 3 year subscription, English

Table 6: OF-Controller HW configuration

- All IBM Hardware components have 12 month support.
- This computer is also connected with two SX-MM fibers to the segregation-switch.
- This Computer is a two unit rack version

6.1.3 Software setup

On the HP segregation switch we are using the new research firmware image of the vendor which supports OF-networking. This SW image is Version 2.02e with firmware version K.14.79 and implements OF 1.0 standard. The NEC access-switches will be used with the installed firmware image in version OS-F3L Ver. 11.1.C.Ae.

On the Rack-Server is a 64-Bit SUSE Linux Enterprise Operating System installed with a Xen-Kernel (2.6.32.27-0.2-xen). This OS has a three year warranty and supports all needed programs like VPN, secure shell and other Linux based packages. Different other guest OS are possible by hosting them with the Xen hypervisor for instance a Debian Linux.

6.1.4 Island Access

TUB workgroups which develop OF experiments will get access to the Control-VLAN of the OF switches and to the IBM-Server to ensure different possibility's of experiments, software testing and software development. The Server can be accessed via SSH.

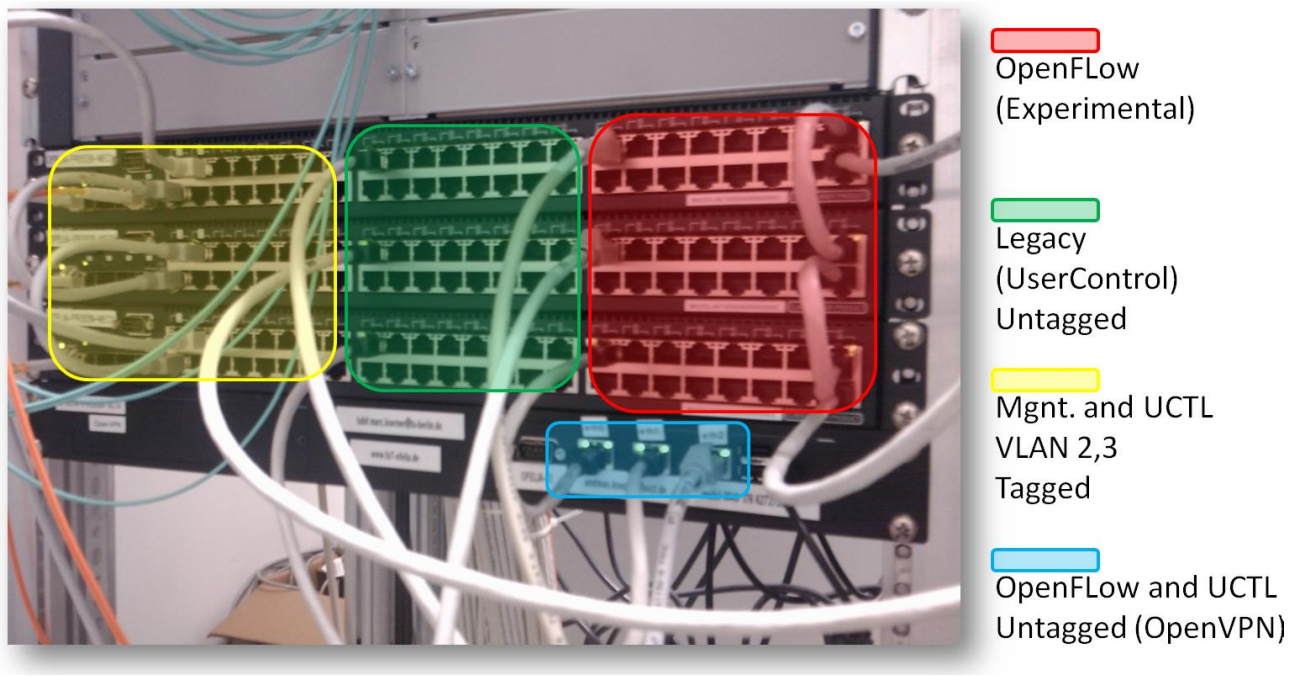

Figure 21: OFELIA equipment in WC FR5539

Additionally dedicated ports are used to give untagged and direct access to the User-Control- and the Experimental-Network. Both networks are connected via an OpenVPN tunnel through the campus network to the Server which is connected with OpenVPN to the IBBT in Gent, too.

Figure 21 shows the distribution of switch-ports with their associated networks.

6.2 TUB island: operational report

During the first phase we had deployed the OpenFlow equipment in the fifth floor of TUB FR building to fulfill the OFELIA project plan for TUB. After a few month without finding researchers to connect to and due to the bad environment conditions in the wiring-centers we decide to relocate the equipment before it takes damage. Unfortunately the situation was not that easy to find a new place and so we use the switches meanwhile for testing purposes to deliver input for the OFELIA architectural discussion. Altogether this causes an island downtime of 3 weeks and we are still waiting to get the switches back from our architectural team.

Thus we have deployed the equipment in a wiring-center at EN building in a temporary setup. In this room we have a very good air-condition and two floors above we had found some researches to connect. Currently we use one of the OpenFlow capable NEC switches in a normal switch mode for aggregation and separating

the different VLAN's. This switch will be replaced with a by other switch and will be added to the experimental OpenFlow setup.

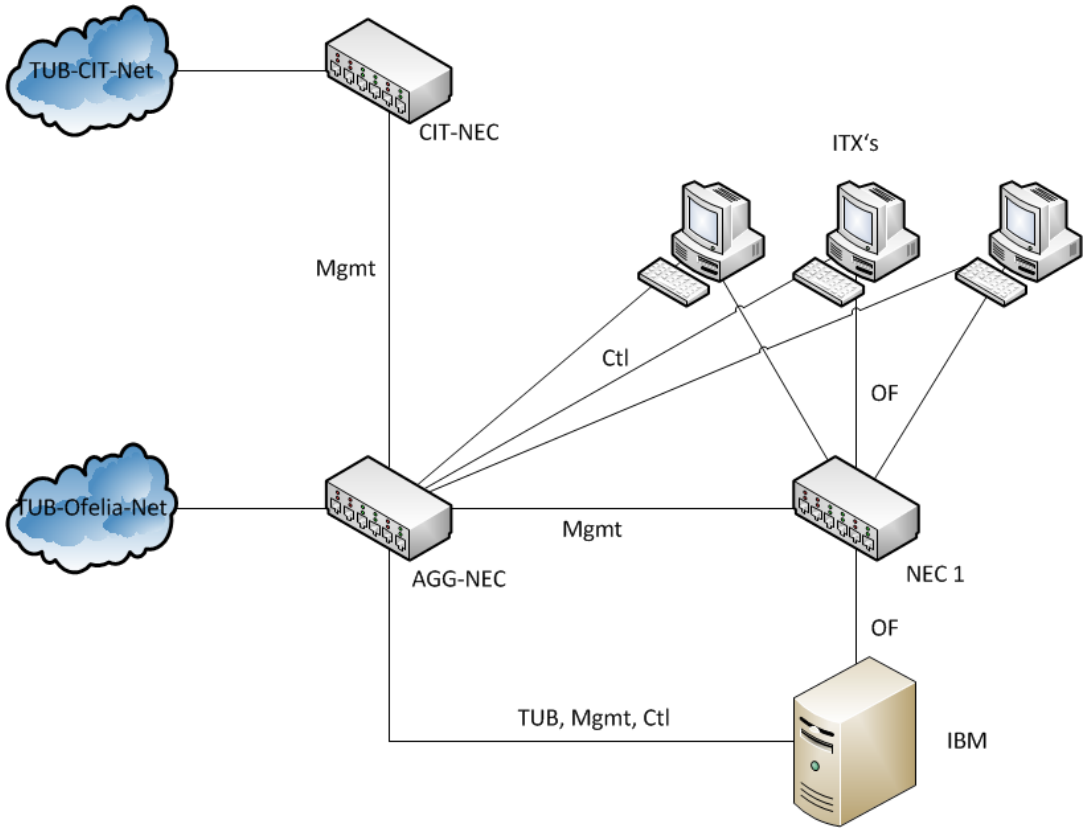

Figure 22: Current testbed topology

The IBM Server, with SLES and Xen, is hosting the control framework, FlowVisor and the OpenVPN service and gateway. All services are encapsulated in VM's. The ITX pc's are using a Debian Squeeze OS and have SSH and VNC access.

Seven researchers of the CIT department use an OpenFlow controlled access-switch for local and Internet connectivity, sliced by a second FlowVisor instance and currently controlled by a NOX with the switching module. The FlowVisor as well as the NOX are hosted on the IBM server using two more VM's

The current configuration separates the testbed which interconnects the researchers of the CIT department form the core testbed, which is part of the federated islands. The simple reason for this is to get first experiences regarding the stability and usability of the testbed for the daily regular use (production traffic). After achieving confidence regarding those aspects, the two testbeds will be united. After this, the new testbed installation will be in line with the original installation (described in 6.1) besides the fact, that all core components are located in the EN instead of the FR building. The FR building will be still connected, especially because of the planned extensions in phase II (e.g. BOWL integration, see 6.3).

For providing a better service for external OFELIA users we decide to make some major improvements of our equipment or rather experimental facility. We have ordered two more servers with 6 Ethernet interfaces for hosting user-vm's. Furthermore we plan to restructure the current switch topology to get an ideal setup for OpenFlow experiments.

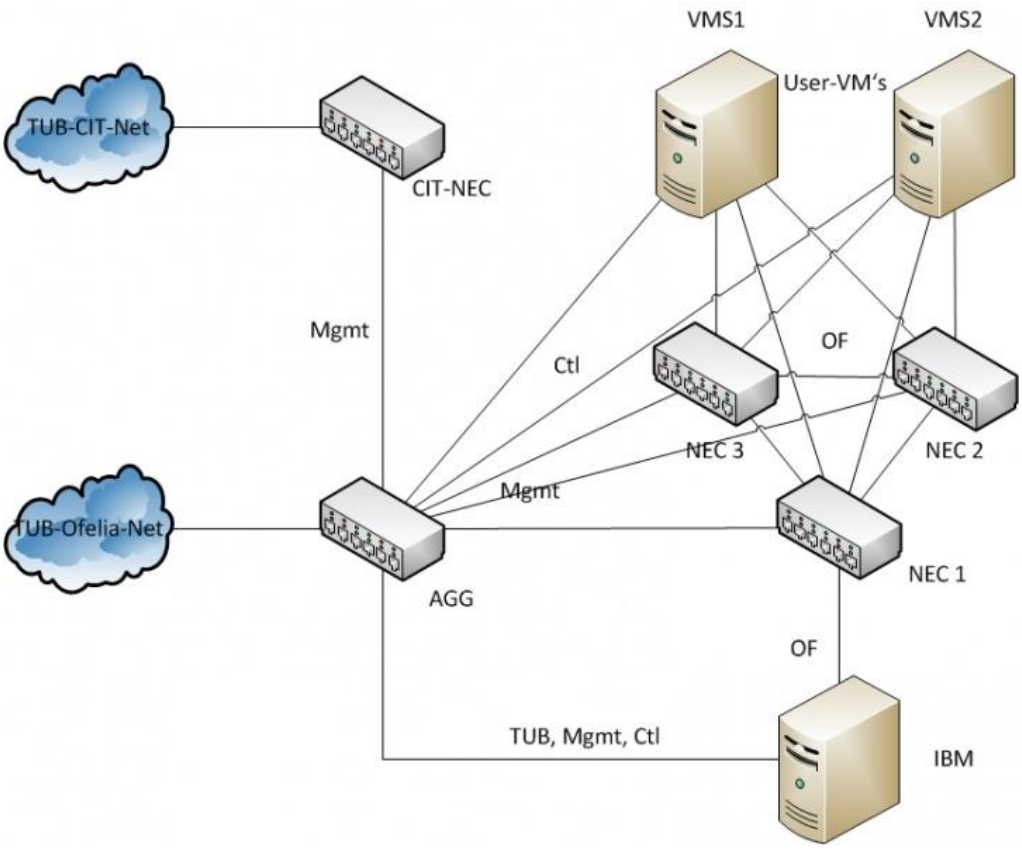

Figure 23: Planned improvements

The 2 additional User-VM servers (named VMS1 and VMS2 in figure 28) are running Debian Squeeze. They are connected to the Ctl and the OF network with dedicated interfaces (eth1-3 = OF; eth4 = ctl).

Number:	Product:
1	Intel Xeon Processor E3-1240 (3,30 GHz, 8 MB)
4	Kingston DIMM 4 GB PC3-10600 ECC
2	1000 GB Western Digital RE4 (7200 UPM/S-ATA II)
1	SuperMicro X9SCM-F; SuperMicro System Management Card (on Board);
2	Gigabit-LAN adapter, Cat5e, RJ-45 (onboard)
1	Intel Gigabit ET Quad Port Server Adapter
1	SeaSonic 400 Watt power supply

Table 7: HW configuration of each of the 2 server providing User-VM's

6.3 TUB island: plans for Phase II

6.3.1 Measurement facility (based on IXIA test system)

The TUB island will be extended by a measurement facility, based on an IXIA test system. Based on a collaboration of TUB with the EANTC AG (European Advanced Test Center, www.eantc.com) and in agreement with IXIA a free loan of an IXIA 1600T test system will enhance the TUB island.

At the beginning, the IXIA 1600T will be equipped with the following interfaces:

- Interfaces fully supported by IxOS and IXNetwork
    - 2 modules with 3 GB Ethernet interfaces each

- 3 modules with 8 Fast Ethernet interfaces each

- Interfaces (older ones, but still very powerful) supported only by IxOS
  - 4 modules with 8 Fast Ethernet interfaces each

Based on the collaboration we expect that additional interfaces can be provided on demand for limited time periods.

The test system provides highly professional test possibilities. The following list describes just a limited selection of the testing features:

- QuickTests for industry--standard test methodologies, including RFCs 2544,2889 and 3918
- Easy-to--use protocol and traffic wizards emulate complex networks with the widest selection of L2 and L3 protocols and traffic profiles.
- Advanced measurements, including four latency modes, delay variation (jitter), inter- arrival time, sequence checking, misdirected packets, packet loss duration, and Rx rates.
- Statistics viewer providing summary level, group level, and per-flow statistics with sophisticated results filtering.
- Multi-field ingress tracking to track flows based on multiple user--defined fields.
- Multi-egress tracking provides a unique ingress/egress results view that compares sent traffic with received traffic, so as to accurately verify what DUT packet field modifications. QoS remarking is easily verified.
- Powerful reporting with one-click report generation, custom report and template builds, customizable charts and graphs, in PDF or HTML formats.

The test system provides a development environment supporting TCL-based development of own test suites.

TUB will integrate this measurement equipment in the TUB island, use it for own measurement purposes, but also offer all OFELIA partners to use it on demand. Therefore, initially a secure access to the test system will be available via VNC. Based on demand, the access possibilities may be extended (e.g. by use of a terminal server).

6.3.2 **BOWL wireless test facility**

6.3.2.1 What is BOWL

BOWLs primary usage is the support of direct Internet access across campus for TU staff. During the integration process with OFELIA it must be recognized that its primary role as an access network must be maintained. Any usage of the facility for the purposes of experimentation must be possible without jeopardizing the primary user. This leads to considerations in terms of management and resource segmentation and protection. These issues will need to be addressed carefully as the integration process progresses.

6.3.2.2 BOWL infrastructure and technical capabilities

The BOWL wireless facility at TU-Berlin is based on 802.11 technology. It is distributed throughout the campus with 46 APs and approximately 100 wireless mesh interfaces. Approximately 50 wireless access interfaces are deployed supporting 802.11/a/b/g modes with a further 12 nodes supporting 802.11n. The nodes can support either 2 or 4 embedded wireless interfaces dependent on the hardware platform in question.

Each of the wireless interfaces can be configured to support multiple virtual APs, thereby offering a highly flexible and extensible virtual substrate. The nodes are connected to the wired infrastructure through a single dedicated Ethernet port running at 100Mb/s. All traffic to and from the wireless network is routed towards the public Internet through a router owned and operated by Tu-BIT, located on the 15th floor of the TEL building, Ernst-Reuter-Platz.

6.3.2.3 BOWL and OpenFlow

The operating system deployed on the BOWL network is an embedded Linux version called OpenWRT. OpenFlow has been ported onto the OS and currently supports the entire feature set of OpenFlow 1.0. Both in-band and out-of-band management is possible for OpenFlow control traffic and a mixture of both may be required for the integration process. The L2switching functionality of Linux is replaced within OpenFlow enabled APs using a port of OpenVSwitch 1.2.1 and in combination with the wireless hardware maximum throughout in modes 802.11a/b/g is possible. Embedded instrumentation is also supported on the BOWL nodes using sFlow a very useful feature for collection real time traffic statistics, thresholding and similar functionality.

6.3.2.4 Integration of BOWL with OFELIA

The hardware limitation of a single Ethernet uplink from the BOWL APs represents a technical challenge for the integration process. In the first case user traffic, both OFELIA generated and also traffic generated by the primary users can potentially interfere with control traffic that must share the same link. Possible solutions to support isolation of the traffic types involve using VLANs. With support of multiple VLANs within the Tub IT network is would be possible to ensure control traffic and user traffic is separated at least in terms of visibility. From the perspective of resource allocation this still remains and unanswered question and is dependent on the capabilities and policies of TuBIT and the BOWL nodes themselves with regard to rate control and its application to specific traffic types. In a similar manner further, potentially more flexible options may be pursued using 802.1ad (QinQ) technology, to allow many VLANs for both experimental, control and primary user traffic, again being pursuant on the offerings from TuBIT. There is also support for tunneling technologies within OpenWRT, for example GRE and OpenVPN. These could also be used to virtualize the access and control environments. In order to implement the OFELIA framework on OpenWRT it shall be accepted that only a single data path is available. Primary User traffic should probably therefore be maintained using either policy within the OFELIA flow space, enforced by FlowVisor, or through static flow entries maintained directly on the APs themselves. The benefits/drawback of both approaches needs to be further investigated. Each AP could potentially host a dedicated controller, in this might be useful for integration into the OFELIA framework. However this approach is atypical and again would require further investigation. It is however imperative that at least 1 of the embedded radios within each AP remains dedicated to the Eduroam network.

7 UEssex Island

7.1 UEssex island: current Status

In this section we describe the current status of UEssex Island setup. The University of Essex Island consists of Layer 2 NEC switches, cluster of physical servers to run virtual machines, traffic generators acting as traffic sources. UEssex OpenFlow network is currently managed using the SNAC controller and FlowVisor is used to create network slices of the underlying network infrastructure.

7.1.1 Topology and connectivity

The topology is separated into the physical structure and the logical connectivity. After functionality validation of the facility topologies real user traffic (researcher and student) should be handled by the OF infrastructure. At the moment only two end-hosts are connected for the validation process.

7.1.1.1 Physical connectivity

The topology at University of Essex comprises of OpenFlow enabled NEC switches and Physical servers for running Virtual machines and OFELIA control framework.

University of Essex has a very good connectivity to the GEANT and JANET networks which connects University of Essex to a wide range of Universities across the world and within the UK. Fig 1 shows the UEssex topology for the phase I of the OFELIA project.

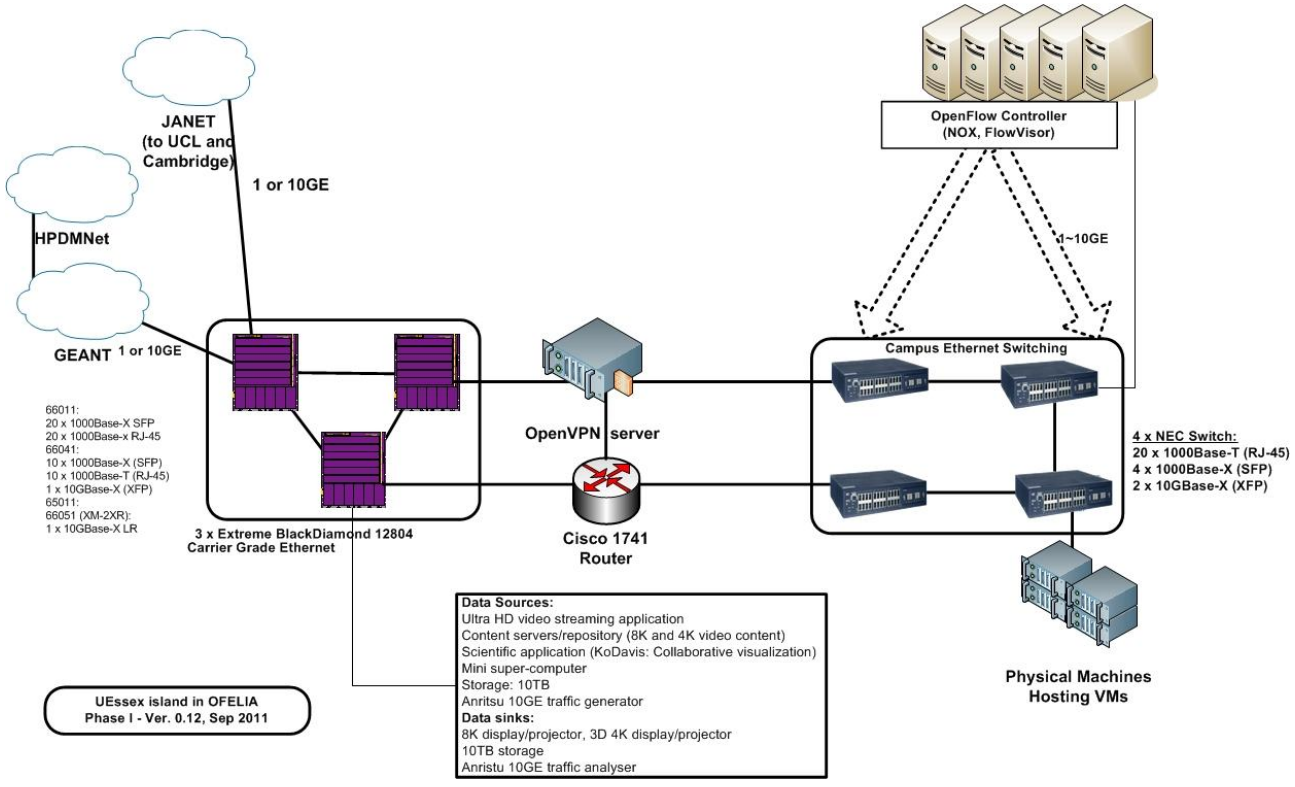

Figure 24: Physical Topology of UEssex OpenFlow island

The 10GE connectivity from GEANT and JANET are terminated on the carrier grade Black Diamond 12804R Switches. An OpenVPN server is connected to the Extreme BD switches using 1GE interface. The external interface resides on the University's network to establish OpenVPN tunnel with the IBBT (Hub of OFELIA infrastructure). A Cisco router sits in between the OpenVPN server and the internal network to route the traffic from the servers to other island's network and route Internet traffic from servers through the University's network. The physical servers are connected to the NEC switches using the 1 GE interfaces. The Physical servers are used for hosting virtual machines for experimental purposes. UEssex island can be

accessed via connectivity through GEANT network or over VPN through the Internet which will be discussed later in the document.

7.1.1.2 Logical Topology

The physical topology shown in Figure 24 is logically separated into Legacy configuration and OF enabled connections in the NEC switches. The legacy part is where the OpenFlow control network is configured. The control network used in the current network is purely out-of-band. The management traffic in the OpenFlow network currently resides on the University's network and it not routed between the OFELIA islands. The OF part of the switch is configured using the Virtual Switch Instance (VSI). The current slicing mechanism used at UEssex is based on VLANs however it might be changed in the future if other slicing mechanisms prove to be more beneficial (i.e., MAC slicing). All the components of the OFELIA control Framework Software (Expedient, Opt-in, FlowVisor and the OF controllers) resides on machines that are part of the legacy configuration in the switch in a dedicated VLAN. The experimental interface from the VMs is directly connected to the NEC switches for experimental traffic. A Calient black Diamond switch is also part of the network test bed interfacing with the NEC switches. But it will not be part of the OpenFlow test bed offered in Phase I of the project.

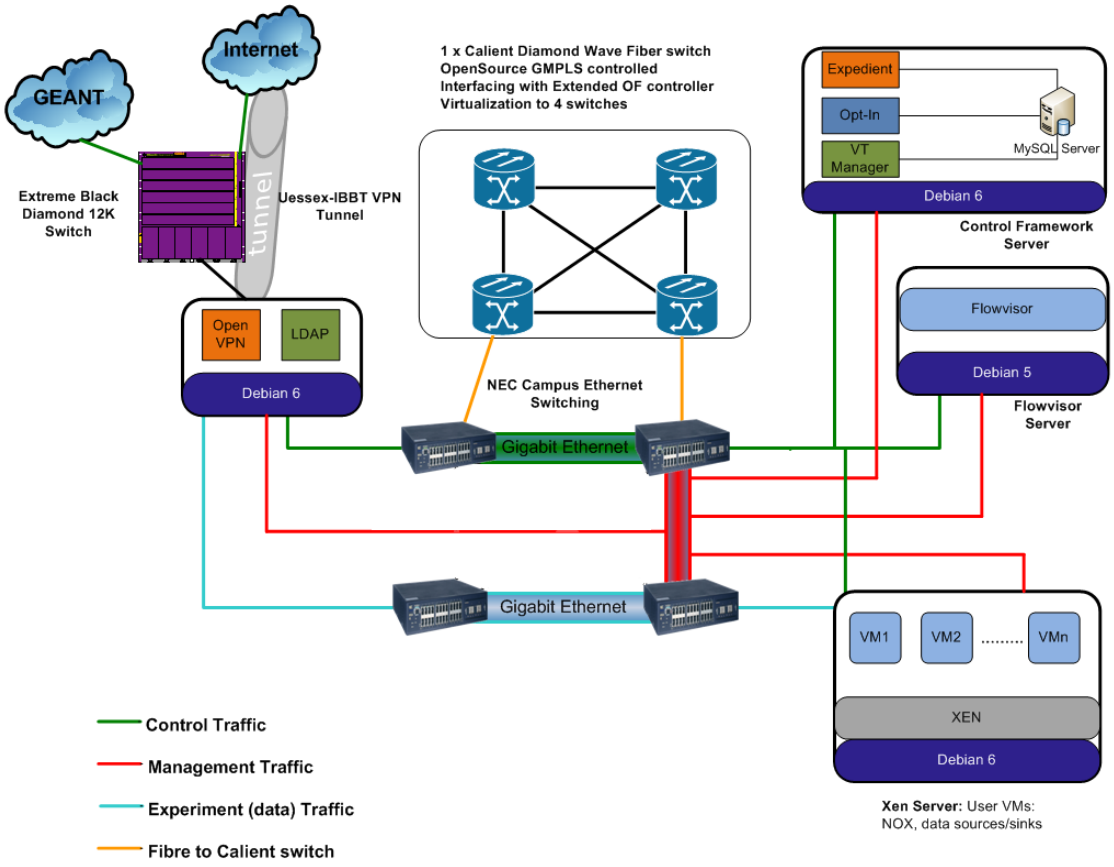

Figure 25: Logical Topology

Figure 25 gives an overview about the current connection status of UEssex island.

Island parameters in facts:

- Management Network
  - Dedicated connection between NEC switches and Extreme Switches
  - NEC switch internal VLAN id 350
  - IP subnet range 155.245.64.0 / 23 (will be moved to 10.216.148.0/22 network)
  - Switch management (Within UEssex Island)

- User Control Network
  - OpenVPN tunneled
  - NEC switch internal VLAN id 2
  - IP subnet range 10.216.20.0 /22
- Experimental Network
  - OpenVPN tunneled

Dedicated NEC switch wiring

7.1.1.3 Inventory

Inventory of OpenFlow switches in the testbed			
Manufacturer	**Model**	**Datapath ID**	**Status**
NEC	NEC IP8800/S3640-24T2XW	0000000000000001	Up and running
NEC	NEC IP8800/S3640-24T2XW	0000000000000002	Up and running
NEC	NEC IP8800/S3640-24T2XW	0000000000000003	Up and running
NEC	NEC IP8800/S3640-24T2XW	0000000000000004	Up and running

Inventory of servers machines in the testbed				
Host name	**Operating System**	**RAM**	**Duties**	**Status**
cseedurham	Debian Squeeze 64-bit Release:6.0.1	8 GB	XEN Server, VMs for advisor,LDAP	Up and running
cseedelphi	Debian Squeeze 64-bit Release:6.0.1	8 GB	XEN Server, ZenOSS	Up and running
cseeopctrl	Debian Squeeze 64-bit Release:6.0.1	2 GB	Control Framework	Up and running

7.1.2 Hardware configuration and setup

In Phase I of the OFELIA project, UEssex Island will be having four NEC IP8800/S3640-24T2XW's, a cluster of physical servers for hosting different applications.

The server hardware configuration used for deploying Control Framework, consisting of Xen Virtual machines, is listed below:

- 4 OpenFlow switches: NEC IP8800/S3640, 24 ports (1/10GE)
- 4 Dedicated Physical servers
  - 2 Dell Power Edge 1950, Intel Xeon(R) Quad Core X5355 @ 2.66GHz, 8GB RAM, 400GB disk (1u)

- 1 Dell Power Edge R200, Intel Xeon(R) Dual Core E3110 @3.00 GHz, 4GB RAM, 600GB disk (1u)
- 1 Dell Power Edge 860, Intel Xeon(R) Dual Core 3060 @2.4GHz, 2GB RAM, 200GB Disk(1u)

The Physical servers are connected to the NEC switches and Carrier Grade Black Diamond 12804R switches using 1GE interfaces

The hardware for end hosts which are virtual machines are deployed on Dell Power Edge 1950 servers. Currently there are only two physical servers for VMs; more powerful servers will be added to the UEssex OpenFlow island at a later stage of the project. In addition to the virtual machines for the end users, there are dedicated server boxes used for UHD applications coders and decoders.

There are dedicated servers for running OpenVPN gateway and Island administrators OpenFlow controller to isolate it completely from the experimental setup as they form the core for interconnection between Islands and operations within the Island. Control Framework (Expedient, Aggregate, Opt-IN managers) are installed on dedicated machine for the first Phase of the project.

7.1.3 Software setup

SNAC version 0.4 will be used as the OpenFlow administrator's controller at UEssex island which includes a policy manager for policy management. The SNAC OpenFlow controller is installed on Debian Lenny Operating system as the source image available from Stanford is based on Debian Lenny. FlowVisor software which is used to slice the network is also installed on Debian Lenny. There are regular patches being applied to the OpenFlow tools and new releases are emerging as OpenFlow is still an emerging technology. The current version of FlowVisor available is version 0.8 which will be used for the phase I of the project. The aggregate managers will be built on Debian Operating systems as the source are built based on Debian which aids in the development process by eliminating the compatibility issues.

The virtual machines environment which is used as the end hosts is based on XEN virtualization which is created using the OFELIA control framework. XEN hypervisor is run on top of the Debian operating system version 6 (Squeeze).

Open LDAP will be used for authenticating external as well as the internal users and also for the VPN connectivity. From UEssex perspective, the LDAP configuration would be a replication of the one used at IBBT with some customization to suit UEssex environment. Since IBBT is the central hub of connectivity between islands and for external users, it is decided to use the same versions and platform for installing Open LDAP so that LDAP configuration and the database remain synchronized with all the other islands.

7.1.4 Island Access

Inter-Island Connectivity

UEssex Island is connected to GEANT Network with 10GE connectivity. There are two ways of connecting to UEssex Island.

- VPN over Internet
- GEANT Connectivity
- On the Islands where connectivity to GEANT network is already established, those Islands can connect to UEssex Island via GEANT through Layer 2 connectivity. In case of Islands currently not connected to GEANT network, connectivity will is established via VPN over the Internet.

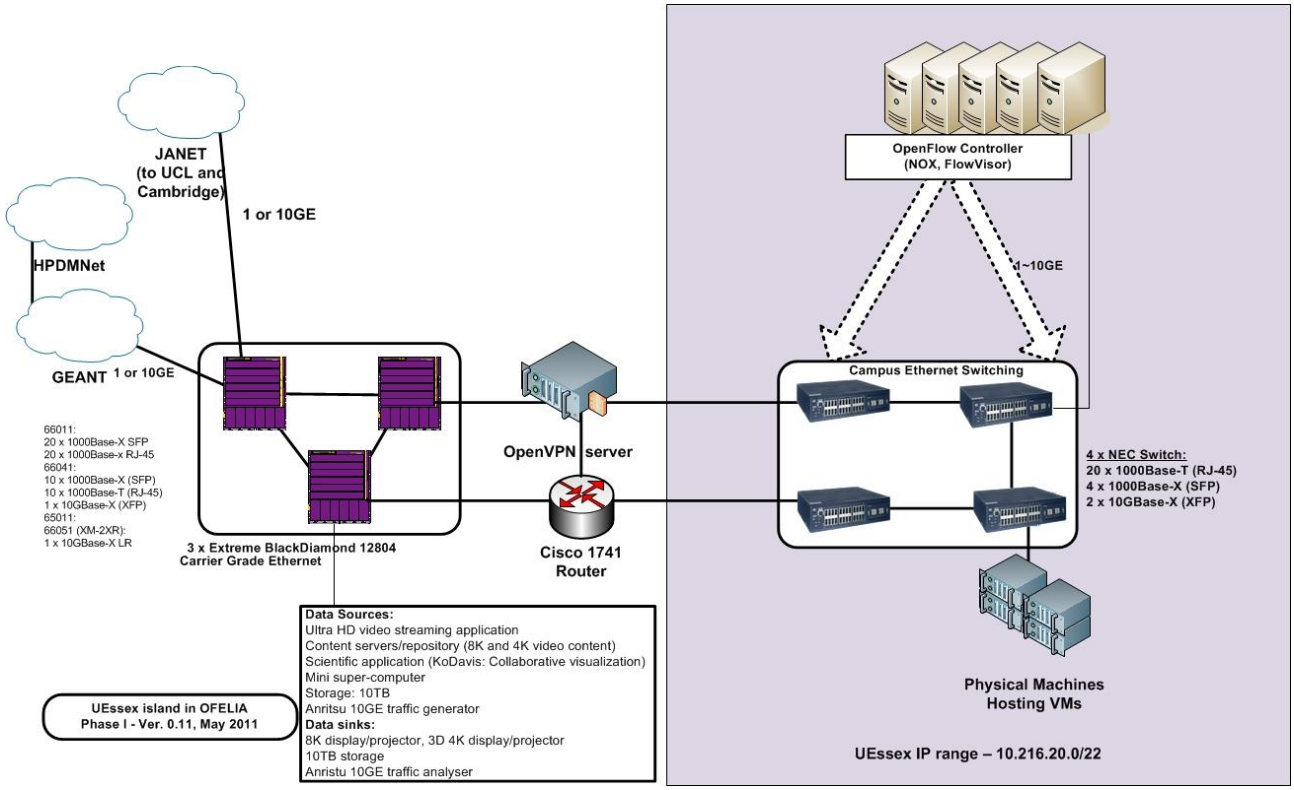

Figure 26: External Connectivity

UEssex will connect to different Islands via IBBT which acts as the Hub site of the OFELIA VPN network. OpenVPN Gateway is setup at UEssex to establish a site to site VPN between UEssex and IBBT and thereby establishing connectivity to all the other Islands. The OpenVPN connection is a bridged VPN connection which creates a logical mesh connectivity across all the islands. SSH connectivity is enabled on servers for SSH access authenticated against LDAP server through certificates or user credentials.

Connectivity for External Users

All the external users are expected to establish VPN connectivity to the Hub VPN site at IBBT. After successfully establishing the VPN connectivity, they will be able to access the reserved OFELIA resources at UEssex island as any other Island users' access.

7.2 UEssex island: operational report

The following operation issues have been encountered in the UEssex island.

- **Issue: User requesting for password reset receives a blank webpage.**
  - **Component:** Expedient
  - **Description:** While testing the CF we observed that when the user requests for a username reset, a web link is sent to his/her email ID which is the expedient link to change the user password. The web link cannot be accessed and hence results in a no access page.
  - **Solution:** To solve this issue, make sure the "SITE_DOMAIN" parameter in localsettings.py is not the hostname but the web link of the expedient. E.g.: SITE_DOMAIN = "exp.uessex.fp7-ofelia.eu"
- **Issue: Identical Virtual Machine names on the same Xen Server leads to discrepancies in the VM access.**
  - **Component:** Expedient
  - **Description:** When 2 or more virtual machines with the same name on the same XEN Server are created in expedient, only one of them can be started. Though both the VMs are created only the first started VM will be accessible.
  - **Solution:** No possible solution found, may require a bug fix.
- **Issue: phpMyAdmin webgui fails**

- **Component:** Control Framework
- **Description:** after installing CF if the phpmyadmin webgui doesn't open then it's probably that your missing the linking package between apache & php.
- **Solution:** check to see if libapache-mod-php5 is installed and if not present install it to get back phpmyadmin webpage access.

- **Issue : VM interfaces are down or if the VM has no access to LDAP**
  - **Component:** Xen physical Server
  - **Description:** If the CF created Xen VMs have no interfaces or if they can't reach LDAP then there is something wrong in the configuration of the Xen server.
  - **Possible Solution:** Steps for debug are as follows:
    - First check if the bridge interfaces are up and they can reach LDAP
    - Check server log and VM configuration file i.e. <vm_name>.conf file
    - Make sure the gateway of the island is reachable and the IP ranges are correct
    - Check if the user, who created the project, is registered in the LDAP

7.3 UEssex island: plans for Phase II

The topology of the UEssex's Island as planned for phase 2 of OFELIA is depicted below (Figure 27):

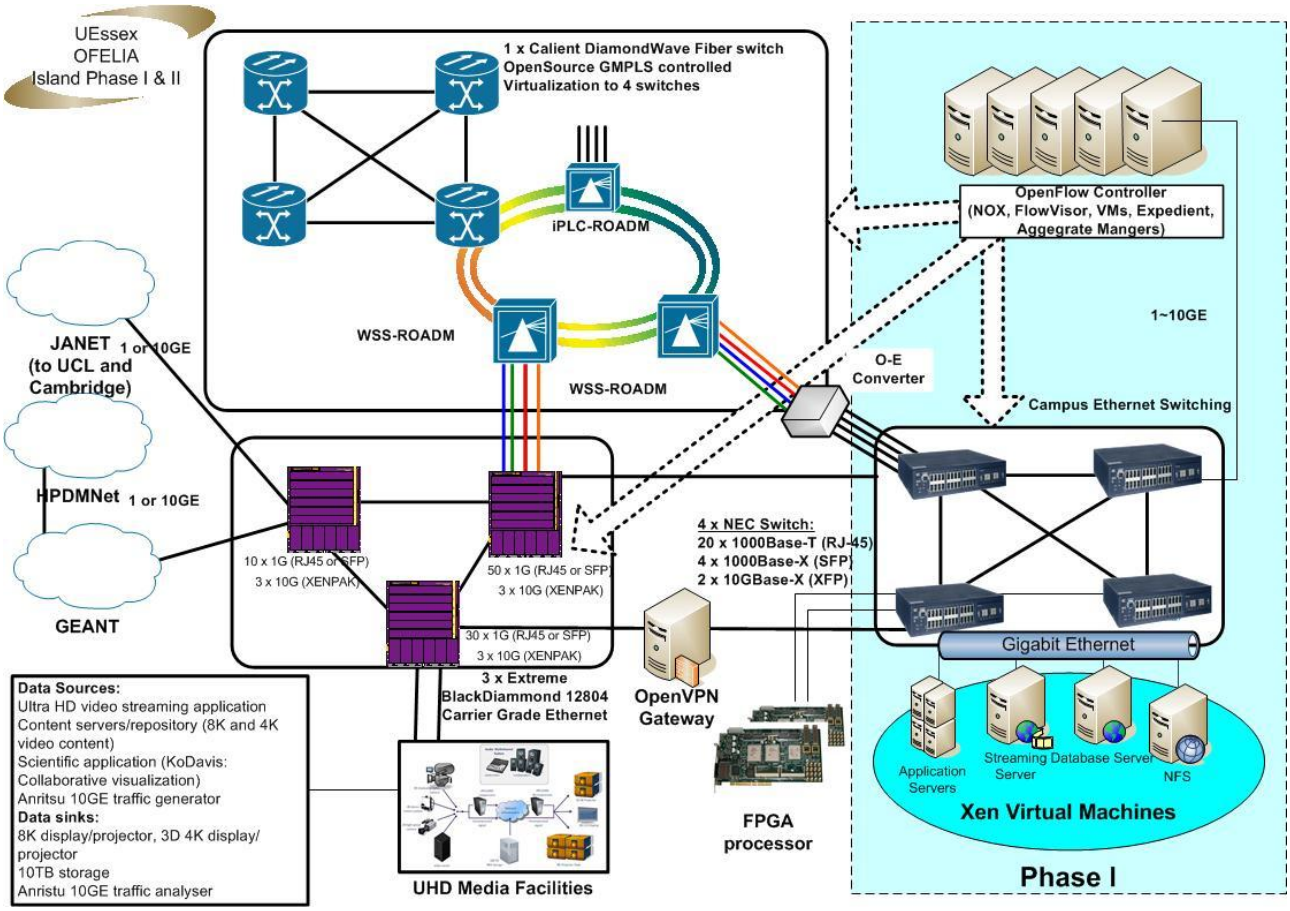

Figure 27: UEssex Island (Phase 2)

7.3.1 Phase 2 Plan

The added building blocks and component in this phase are:

1. 3 x Carrier Grade Ethernet switches (Extreme BlackDiamond 12804) (Will be OpenFlow enabled in late 2011)
2. Extra OpenFlow controllers

3. 2 x Virtex-4 FPGA boards with 1GE network interface
4. 3 x ADVA FSP 3000 DWDM ROADM nodes
5. UEssex-i2CAT connectivity via GEANT (for connectivity test via GEANT)
6. UHD Media Facilities for Source & Sink

7.3.2 Phase 2 Description

1. Carrier Grade(CG) Ethernet switches (Extreme BlackDiamond 12804):
   - Connectivity to Extreme BlackDiamond Switches is established, but the new OpenFlow firmware image is yet to be delivered by Extreme networks. We are expecting to receive the OpenFlow firmware image by early October 2011.
2. Extra OpenFlow controllers
   - Extended OpenFlow controller for optical domain has been developed which interacts with Calient DiamondWave fiber switches.
   - To develop/extend the extended OpenFlow controller for optical nodes and in particiular to interact with ADVA ROADMs.
3. Virtex-4 FPGA boards with 1GE network interface
4. ADVA FSP 3000 DWDM ROADM nodes
   - Currently we have 2 milestones targeted, which will integrate ADVA optical equipments into the OFELIA facility. Milestones cover enabling OpenFlow in optical equipments and also to build an extended optical OpenFlow controller.
   - The envisioned topology of the ADVA optical equipments is as follows:

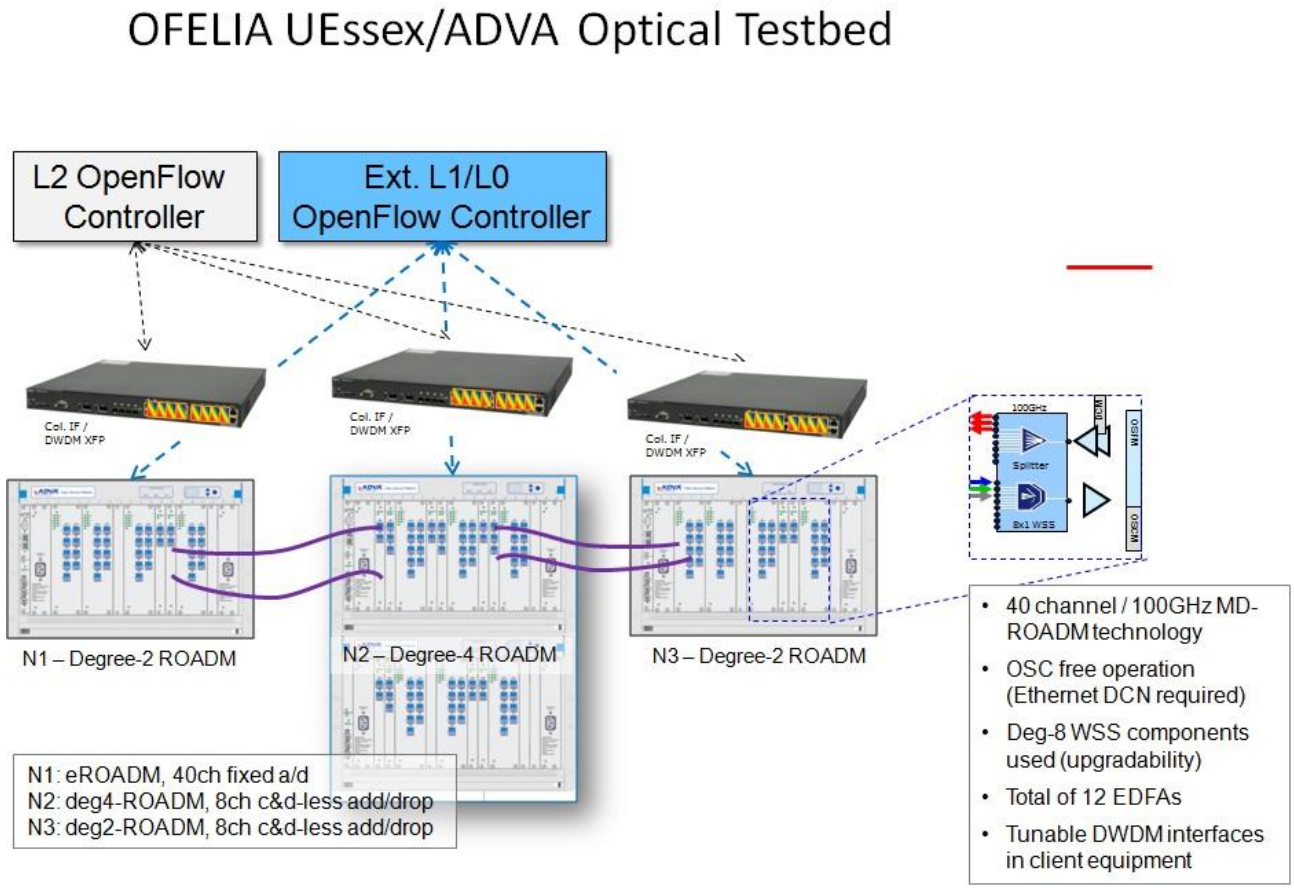

Figure 28: ADVA optical equipment topology

5. UEssex-i2CAT connectivity via GEANT
    - The Carrier Grade (CG) Ethernet switches will provide the connectivity of this island to the OFELIA hub in IBBT (via GEANT, 10GE). There is no particular limitation for this connectivity and all addressing schemes, VLAN (Q-in-Q), PBT, PBB is supported for the connectivity. Since i2CAT has also the existing connectivity to GEANT the plan is to have this connectivity as an example use-case for other partners and also to identify potential issues and considerations for this intra-island connectivity.
6. UHD & Traffic Generation Faculties:
    - The data sources are:
        i. Ultra high-definition video streaming application
        ii. Content servers/repository with 8k and 4k video content
        iii. Scientific application (KoDavis: Collaborative Visualization of Huge Atmospheric Data)
        iv. 10TB storage
        v. Anritsu 10GE multi-layer, multi-protocol traffic generator
    - Data sinks:
        i. 8k display/projector, 3D 4K display/projector
        ii. 10TB storage
        iii. Anritsu 10GE multi-layer, multi-protocol traffic analyzer (which is capable to analyze network from layer 2 upward with a very high granularity on traffic filtering and monitoring)

8 EICT island

Within project OFELIA EICT operates a specific island for testing and development of the OFELIA control framework. At the end of phase I of project OFELIA, regular research operation will start in the production islands. In order to support further development and testing, the EICT island hosts unstable branches of the control framework and allows testing of new features or test necessary bug fixes. Note, the EICT island is not meant for production use (with a single exception currently, see below). It has been assigned the IP address range 10.216.24.0/22. The EICT island consists of the following components:

An island gateway for interconnecting to OFELIA's central hub located at IBBT.

The island gateway is connected to the public Internet via the Deutsches Forschungs-Netzwerk (DFN/German Research Network). It hosts endpoints for terminating VPN connections to OFELIA's central hub and other islands. The EICT island is fully connected, so that experimental traffic as well as the control network is connected to the remaining infrastructure. The island gateway runs currently with Debian 6.0 and is capable of hosting various virtual machines using a XEN bare metal hyper visor. These virtual machines are used for deploying different variants of OFELIA's control framework: at present the stable and the unstable branch of the control framework's GIT repository are deployed. It is planned to provide for each published version of OFELIA's CF an installation for testing in case of occurrence of any severe problems. EICT's island gateway provides four 1 GbE Ethernet interfaces, from which three are currently in use: while the first serves as uplink to the public Internet, the latter two carry control and experimental traffic, respectively. The island gateway also implements OpenVPN access for testing VPN based user access.

Two IBM servers type x3550 with a Xeon dual core CPU and 12GB of memory for hosting virtual machines.

These machines are intended for hosting user defined virtual machines that convey OpenFlow controller instances and/or act as data sinks/sources for users. However, these servers are not intended to host a large amount of virtual machines as the number of available CPU cores is rather limited. Both machines carry the OFELIA default physical node configuration based on Debian 6.0 and use a XEN based bare metal hyper visor. Following OFELIA's current architecture, the Linux kernel bridging subsystem is used for forwarding experimental traffic from and to the VMs to the connected OpenFlow switching device. Each node provides two 1Gb Ethernet interfaces for control and experimental traffic.

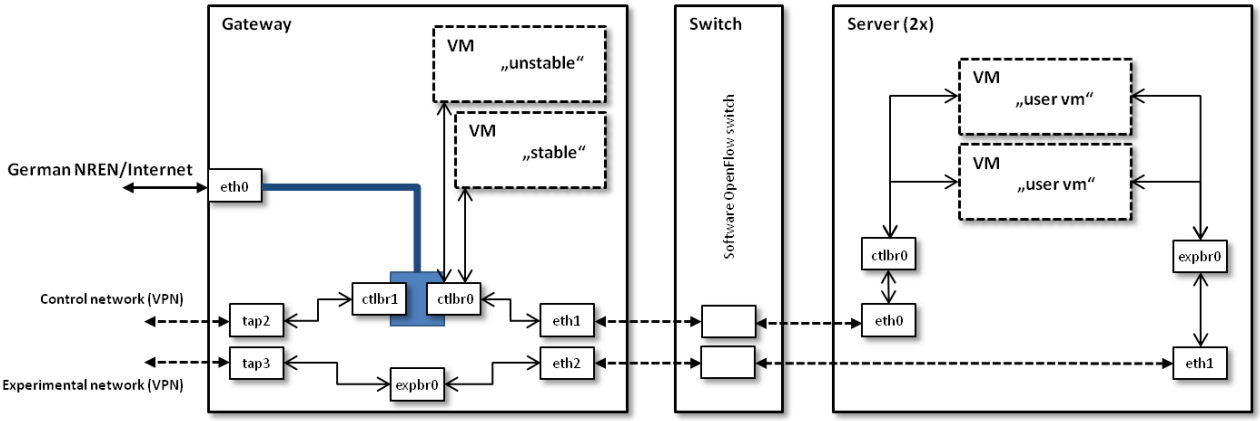

Figure 29 - EICT test and development island

A set of virtual machines hosted by the island gateway carrying different releases of the OFELIA control framework: “stable”, “unstable”...

The “stable branch” virtual machine deployed in the island gateway provides the user registration web application and thus implements the first contact point for users with the facility. This functionality is part of the production system in OFELIA, i.e. user registrations are added to the LDAP infrastructure deployed in the facility from this VM instance. Besides user registration, this VM also hosts the FlowVisor instance for the EICT island. At present, an “unstable branch” virtual machine is currently used for testing upcoming developments. The machine is connected to the control network only. Additional machines hosting releases and the head of development are deployed as well.

A single OpenFlow switching device. This may be a Linux PC running openvswitch or any other OpenFlow capable device.

It can be foreseen that OFELIA will adapt the facility to new versions of OpenFlow in the future. Currently, a software based OpenFlow switch is used for connecting the XEN nodes and the island gateway. This offers adequate flexibility to replace the OpenFlow implementation.

9 NEC island

Within OFELIA, NEC has set up an island for testing, development and support of NEC equipment used in the project. This island is project internal and is not visible to the public. It will host stable and unstable versions of the control framework to be able to test new software and firmware versions of tools and switches and help in bug fixing the stable software versions.

9.1 Topology and connectivity

The island gateway is – as the EICT island – connected to the public Internet via the Deutsches Forschungs-Netzwerk (DFN/German Research Network). It hosts endpoints for terminating VPN connections to OFELIA's central hub and other islands. The NEC island is also fully connected. Experimental traffic as well as the control traffic can be exchanged with the other OFELIA islands. It is reachable with the OFELIA network by the 10.216.28.0/22 network.

9.1.1 Topology

The island currently consists of three physical machines and an NEC switch. One machine hosts the island gateway and acts as router and firewall. Its OS is Debian 6.0 and it runs the OpenVPN connectivity to the other island. It already implements an OSPF router to facilitate route distribution throughout the islands and is also prepared to host OpenVPN access facilities for VPN based user access.

The second machine provides all other OFELIA services like LDAP, DNS and Control Framework. It also runs Debian 6.0 as OS. The third machine is used as XEN hypervisor to deploy additional virtual machines when needed and test Virtual Machine distribution functionality in the Control Framework.

Finally a NEC IP8800/S3640 OpenFlow switch with 48 ports is integrated into the NEC island. It will be used to test new firmware versions and to debug configuration and operation issues that occur in other islands using these switches.

9.1.2 Status

The basic infrastructure is ready and the link to the other OFELIA islands is created. Still the Control Framework needs to be fully deployed and tested within the next weeks.

9.2 NEC Island: plans for Phase II

It is planned that the NEC island integrates an OpenFlow switch based on the OpenWrt firmware image to implement OpenFlow for Wireless links. It might also host some lab internal projects that are closely related to OpenFlow and would benefit from such a testing island.

10 CREATE-NET island

The CREATE-NET island is an island contributed by the new partner CREATE-NET that joined the consortium through the first round of open calls. Therefore, the CREATE-NET island was not operational during Phase I and thus no description of the current status and according operational report could be provided, but only the future plans for Phase II.

10.1 CREATE-NET island: plans for Phase II

10.1.1 Topology and connectivity

CREATE-NET island deployment will be performed in two different stages. In the first stage the whole OF Network will be located at CREATE-NET server farm whereas, in a second stage, part of OpenFlow hardware and some server-class machines will be relocated on a geographically distributed experimental facility located in the city of Trento (Italy).

In the following a description of these two stages will be provided in more detail.

10.1.1.1 CREATE-NET deployment: Stage I

The core of the management LAN is composed of two enterprise switches both of them fully interconnected with all OF network elements. Redundancy, reachability and load-balancing policies will be applied on the management LAN.

Internet connectivity will be provided by using a firewalled router. The interconnection with the other OFELIA islands will be obtained through an OpenVPN connection.

Three different VLANs will be used for management purposes:

- VLAN 99 (MANAGEMENT): traffic between Island components as used by Island Admin, and won't be routed on the global network
- VLAN 999 (CONTROL): OpenFlow Protocol traffic between VeRTIGO network virtualization layer[1], OpenFlow switches and OpenFlow controllers. Also used for User's SSH connections.
- VLAN 998 (INTERNET): provide Internet connectivity to the devices

Three physical servers will be used to host management servers through VirtualBox VMs. Furthermore, three more servers will host several Xen VMs for user experimentations.

[1] VeRTIGO is the acronym for the network virtualization architecture to be deployed within WP8 which extends the virtualization functionalities provided by FlowVisor (see Sect. 10.1.1.4 for more information)

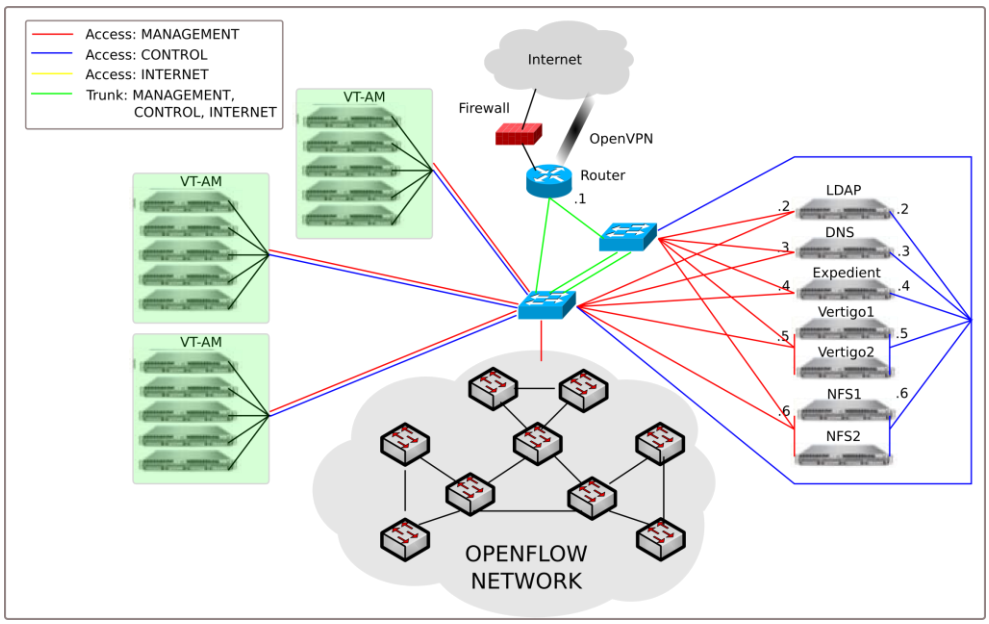

Figure 30: CREATE-NET island topology, stage I

10.1.1.2 CREATE-NET deployment: Stage II

During the second stage, part of the OpenFlow-enabled network nodes will be relocated within a geographically distributed facility located in the city of Trento. Beyond a limited service disruption of CREATE-NET island (that will be communicated well in advance via OFELIA mailing list and to researchers running their experiments over the island), this change won't impact the operations of the island itself that will be accessible as per the previous stage.

The facility is composed of three different locations interconnected through a dedicated fiber pair (max link distance 8.6 Km) in a ring topology as depicted in Figure 31.

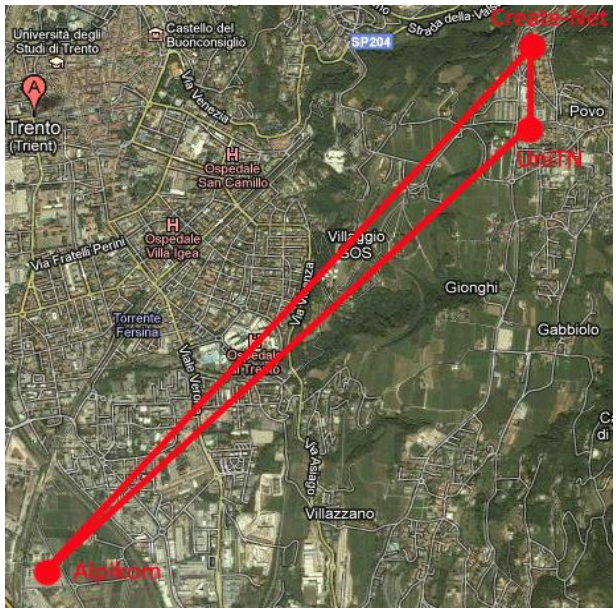

Figure 31: CREATE-NET island geographical network map, stage II

The three locations of the experimental facility are: CREATE-NET, University of Trento (Department of CSEE) and Trentino Network[2], a public regional network operator which is providing Internet connectivity and other telecommunication services to the Public Administration all over the region. The idea behind this is to open up the island to users located in the local University (e.g. students, other research teams, etc) or within the Public Administration and willing to test novel OpenFlow-based services.

Figure 32 shows the physical topology of the CREATE-NET "distributed" island. As depicted, the island will be divided into three sub-islands each of them located at different locations of the testbed and including three OpenFlow switches and one server-class PC hosting users VMs.

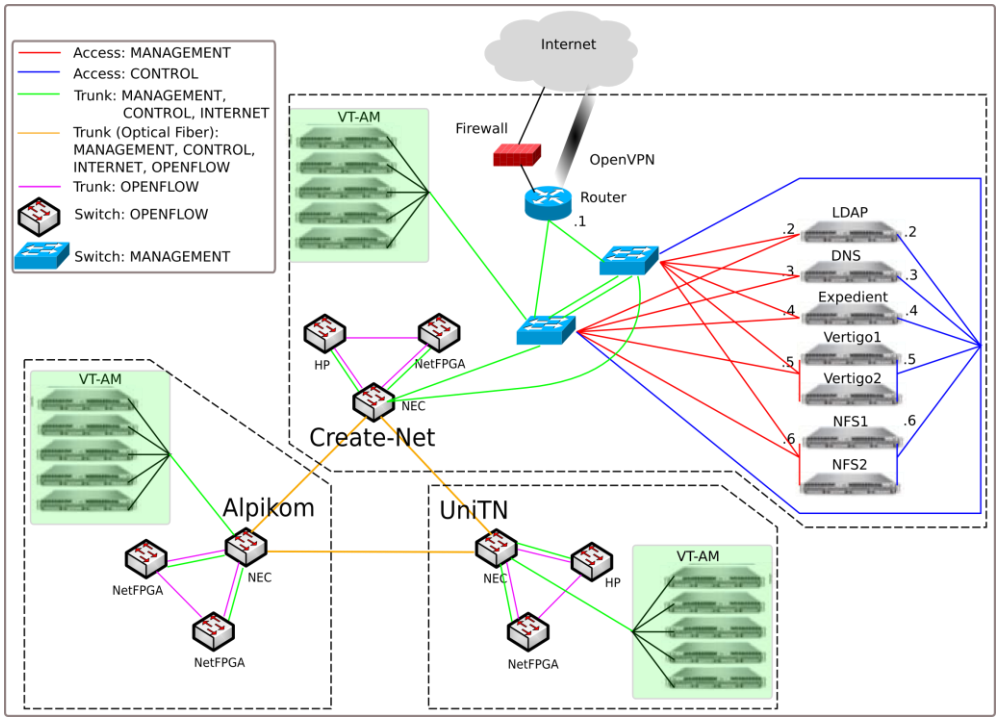

Figure 32: CREATE-NET island topology, stage II

10.1.1.3 IP Addressing Schema

Following the Ofelia's addressing convention, the following ranges will be used:

- **Control network (VLAN 999):** 10.216.32.0/22
- **Management network (VLAN 99):** 10.216.160.0/22
- **Data network (OpenFlow network):** No restriction

[2] Through their server farm hosted in Alpikom (a local ISP) premises.

Hostname	Network	VLAN	IP Address	Netmask
CN-Router	Internet	---	193.206.22.102	/28
	VPN-Control	---	10.216.0.32	/20
	VPN-Management	---	10.216.128.32	/20
	Control	999	10.216.32.1	/22
	Management	99	10.216.160.1	/22
CN-LDAP	Control	999	10.216.32.2	/22
	Management	99	10.216.160.2	/22
CN-DNS	Control	999	10.216.32.3	/22
	Management	99	10.216.160.3	/22
CN-Expedient	Control	999	10.216.32.4	/22
	Management	99	10.216.160.4	/22
CN-Vertigo	Control	999	10.216.32.5	/22
	Management	99	10.216.160.5	/22
CN-Vertigo1	Control	999	10.216.32.7	/22
	Management	99	10.216.160.7	/22
CN-Vertigo2	Control	999	10.216.32.8	/22
	Management	99	10.216.160.8	/22
CN-NFS	Control	999	10.216.32.6	/22
	Management	99	10.216.160.6	/22
CN-NFS1	Control	999	10.216.32.9	/22
	Management	99	10.216.160.9	/22
CN-NFS2	Control	999	10.216.32.10	/22
	Management	99	10.216.160.10	/22

Figure 33: CREATE-NET island, IP addressing schema

10.1.1.4 Slicing

Differently from other OFELIA islands, the CREATE-NET island will use VeRTIGO network virtualization framework (instead of FlowVisor) developed in Work Package 8 to instantiate and manage experimental slices that will be allocated on virtual topologies decoupled from the physical topology of the island.

VeRTIGO will allow the instantiation of generalized virtual topologies in an OpenFlow network through the implementation of virtual links as aggregation of multiple physical links and OpenFlow-enabled switches. These virtual topologies will provide researchers flexibility in designing their experiments with arbitrary network topologies on a given physical infrastructure.

Although the CREATE-NET island will be build around VeRTIGO, it will be compliant with the OFELIA federation requirements. In particular VeRTIGO will support the slicing mechanism defined in the OFELIA architecture.

10.1.2 Hardware configuration and setup

This section describes the hardware (and its configuration) the CREATE-NET island is composed of. The section is divided in two: (i) description of the switches the island is composed of, (ii) description of the servers hosting the virtual machines.

10.1.2.1 OpenFlow Switches

As shown in Table 5, three different kind of OpenFlow-enabled switches will be employed. NEC switches will act both as core-switches and as access-switches, whereas other switches will only act as access-switches. Interconnection between switches will be provided at Gbps bandwidth, using both optical fibers and Ethernet cables.

Number	Product
3	NEC IP8800/S3640-24TWLW with 24 10/100/1000 BASE-T LAN interfaces and 4x SFP
2	HP ProCurve 3500 with 24 port 100 BASE-T LAN interfaces and 2x SFP
4	NetFPGA modules with 4 port 1000 BASE-T LAN interfaces

Table 8: OpenFlow switches

10.1.2.2 Management Switches

To provide reachability, redundancy and load balancing between servers and the OpenFlow network, the management network will be equipped with two Gigabit switches, both interconnected with all devices.

Number	Product
2	Dell PowerConnect 5324 with 24 10/100/1000 BASE-T LAN interfaces and 2x SFP

Table 9: Management switches

10.1.2.3 Servers

The island will be equipped with six server-class machines in order to provide management services (VPN, LDAP, DNS, Ofelia Control Framework, NFS, FlowVisor, and VeRTIGO) and to host VMs used during the experimentations.

Number	Product
2	Dell PowerEdge 1850, with Intel Xeon 4 core proc., 3 GB RAM, 2x73 GBytes RAID1 disks
1	Dell PowerEdge 1750, with Intel Xeon 4 core proc., 2 GB RAM, 2x73 GBytes RAID1 disks
3	Server-class machines, with Intel Xeon 4 core proc., 16GB RAM, 2x500 GBytes RAID1 disk

Table 10: Servers

10.1.3 **Island Access**

10.1.3.1 Inter-island connectivity

Connectivity to other OFELIA islands will be provided through an OpenVPN tunnel with the IBBT island, which acts as the hub site of the OFELIA VPN network. The OpenVPN connection is a bridged VPN connection which creates a logical mesh connectivity across all the islands.

SSH connectivity is enabled on servers for secure access to the island. Authentication is performed against LDAP server through certificates or user credentials.

10.1.3.2 Connectivity for External Users

All the external users are expected to establish VPN connections to the VPN Hub located at IBBT. After successfully establishing the VPN connectivity, they will be able to access the reserved OFELIA resources at the CREATE-NET island as for the other islands.

11 CNIT island

The CNIT island is an island contributed by the new partner CNIT that joined the consortium through the first round of open calls. Therefore, the CNIT island was not operational during Phase I and thus no description of the current status and according operational report could be provided, but only the future plans for Phase II.

11.1 CNIT island: plans for Phase II

The CNIT islands are contributed by the new partner CNIT that joined the consortium through the first round of open calls. Therefore, the CNIT islands were not operational during Phase I and thus no description of the current status and according operational report could be provided, but only the future plans for Phase II.

11.2 CNIT islands: plans for Phase II

There will be two separated islands, one in Rome (CNIT-RM) and one in Catania (CNIT-CT).

The CNIT-RM island will be based on Linux PCs running OpenVSwitch, the CNIT-CT island will be based on OpenFPGA switches.

The two islands will be used to design and implement the Content Centric functionality into the Open Flow architecture. This CCN functionality will also be integrated in the OFELIA control framework.

Moreover, the two islands will fully implement the "standard" OFELIA control framework so that they will be exposed to OFELIA users for any kind of experiments.

11.2.1 Topology and connectivity

11.2.1.1 IP Addressing Schema

The **provisional** decision is to allocate two separate "Island IDs" to the CNIT Rome and Catania island. This decision will need to be confirmed and agreed by OFELIA partners before putting the islands into production. Taking into account this provisional decision and following the Ofelia's addressing convention, the IP address to be used will be:

CNIT-RM island

- **Control network:** 10.216.36.0/22
- **Management network:** 10.216.164.0/22

CNIT-CT island

- **Control network:** 10.216.40.0/22
- **Management network:** 10.216.168.0/22

The topology of the islands is represented in Figure 34 and in Figure 35.

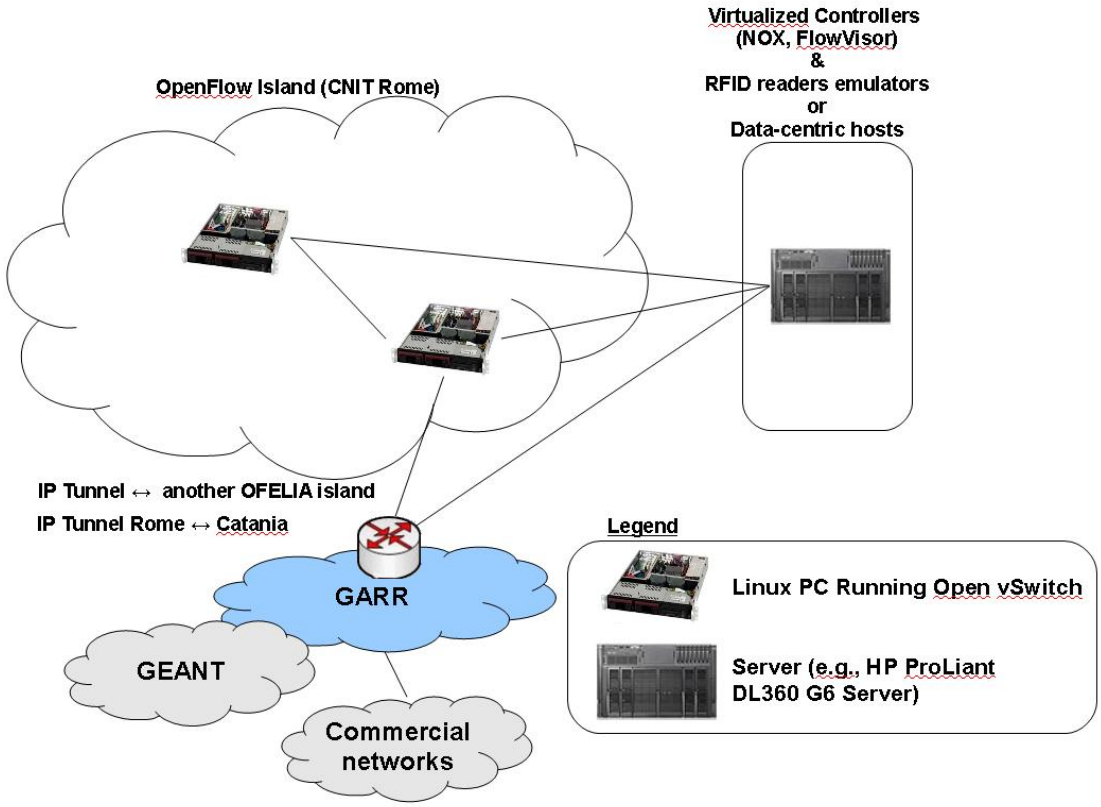

Figure 34: CNIT-RM island topology, stage II

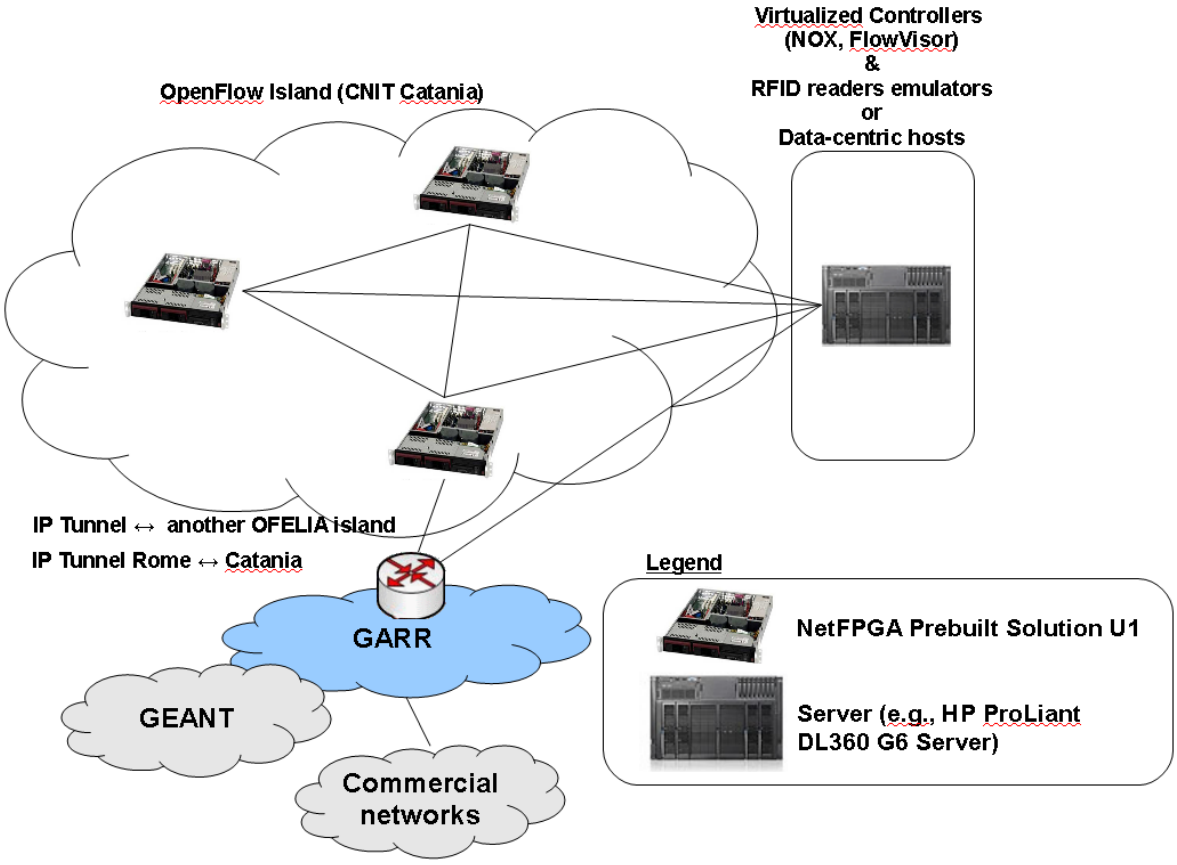

Figure 35: CNIT-CT island topology, stage II

11.2.2 Hardware configuration and setup

11.2.2.1 CNIT-RM island

- **Hardware:** 2 x Linux PC running Open vSwitch, 1 x server, e.g., HP ProLiant DL360 Servers (2x Intel Xeon 5600 series, 16 GB RAM, 685 GB hard drive).
- **Experimental topology:** The two OpenSwitches will be connected with each others. Nevertheless, it will possible for each user to configure whether such connection should be considered active or not. The server will be connected to the two OpenSwitches.
- **OpenFlow Controllers:** Users will be provided with a NOX controller per slice running on VMware virtual machines installed in the server. Although it will be possible to run custom images of NOX on the virtual machines, this will not be enabled as long as there are no reliable technical solutions avoiding detrimental behaviors by NOX installed by third parties. FlowVisor will also run as a virtualized server in the server machine to *slice* the islands network resources.
- **External Access:** Rome CNIT Research Units have 1Gbps connections with GARR, which is the Italian national research & education network. GARR is integrated with the GEANT and is interconnected to several commercial networks such as GX (5 Gbps), TELIA (5 Gbps), MIX (4 Gbps), NAMEX (13 Gbps), TOP-IX (1 Gbps), TIX (1 Gbps), VSIX (1 Gbps). Furthermore, for each of the two CNIT islands two of the OpenFlow switches will be connected to the other CNIT island and to one of the other OFELIA islands through an IP tunnel.

11.2.2.2 CNIT-CT island

- **Hardware:** 3 x NetFPGA Prebuilt Solution 1U - Dual Core (the NetFPGA mounts 4x1Gb Ethernet, the hosting PC mounts 1x1Gb Ethernet), one server, e.g., HP ProLiant DL360 Servers (2x Intel Xeon 5600 series, 16 GB RAM, 685 GB hard drive).
- **Experimental topology:** NetFPGA OpenFlow switches will form a completely meshed topology. Nevertheless, it will possible for each user to configure the desired topology where her experiment should be executed. The server will be connected to all the three NetFPGA hosting PCs.
- **OpenFlow Controllers:** Same as in the CNIT-Rome Island.
- **External Access:** Also Catania CNIT Research Units has 1Gbps connections with GARR. Thus the same description applies.

11.2.3 Island Access

11.2.3.1 Inter-island connectivity

Connectivity to other OFELIA islands will be provided through an OpenVPN tunnel with the IBBT island, which acts as the hub site of the OFELIA VPN network. The OpenVPN connection is a bridged VPN connection which creates a logical mesh connectivity across all the islands.

SSH connectivity is enabled on servers for secure access to the island. Authentication is performed against LDAP server through certificates or user credentials.

11.2.3.2 Connectivity for External Users

All the external users are expected to establish VPN connections to the VPN Hub located at IBBT. After successfully establishing the VPN connectivity, they will be able to access the reserved OFELIA resources at the CREATE-NET island as for the other islands.

11.2.4 Time plans

According to the DoW, we have the milestone **MS92** "Initial key functionality deployed in CNIT islands" foreseen for M18 (Feb 2012). We plan to have hardware available by mid November 2011, and to close the setup of the island and of the "standard" OFELIA Control Framework by December 2012.